

Módszertani füzetek pedagógiai vezetőknek

1. A pedagógiai értékelés



Csapó Benő

*A tanulói teljesítmények
értékelésének
mérési módszerei*

3

XC 18639

Művelődési Minisztérium
Vezetőképző és Továbbképző Intézete
Budapest, I. Szentháromság tér 1-3.

Veszprém Megyei Pedagógiai Intézet
Veszprém, Tolbuhin u. 21.

MÓDSZERTANI FÜZETEK PEDAGÓGIAI
VEZETŐKNEK

I.

A PEDAGÓGIAI ÉRTÉKELÉS

CSAPÓ BENŐ

A TANULÓI TELJESÍTMÉNYEK ÉRTÉKELÉSÉNEK
MÉRÉSES MÓDSZEREI

Sorozatszerkesztő:

Övegesné Terjék Éva
Simon P. Györgyi

1987/88.

SZTE Egyetemi Könyvtár

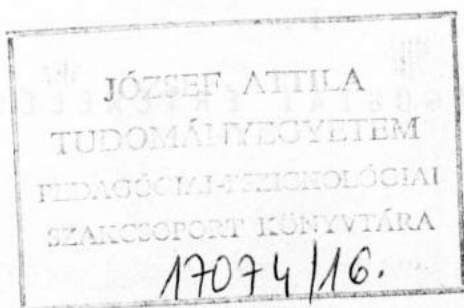


J000689006

Peto. tan. 16 - KA

Tudományos tanácsadó:

Báthory Zoltán



Kiadja: a Művelődési Minisztérium
Vezetőképző és Továbbképző
Intézete és a
Veszprém Megyei Pedagógiai
Intézet

Felelős kiadó: Kapor Károly
Készült: 4.200 példányban a
Veszprém Megyei Pedagógiai
Intézet házi nyomdájában

KF eng.sz.: 56645/1987.

ISSN 0238-2067

ISBN 963 01 8711 6

E L Ő S Z Ó

A hazai újíto, jobbító szándékok, jó törekvések számos esetben szenvedtek kudarcot az új szerepekhez, tartalmakhoz kapcsolódó új módszerek hiányában. Módszertani sorozatunk szerkesztői ennek fontosságára szeretnék olvasók figyelmét irányítani - ezért ajánlják a füzeteket mindazoknak, akik e törekvések jegyében végzik mindennapi munkájukat. Az iskola-, a közoktatásirányítás modellváltásának időszakában tehát a pedagógusoknak, a szakmai munkaközösségek és a nevelő-oktató intézmények vezetőinek, a szaktanácsadóknak és a közoktatás irányítóinak szeretnénk segítséget nyújtani.

Reméljük, hogy sorozatunkkal hozzájárulunk a napi gyakorlat, az önképzés, az intézmények belső továbbképzése, illetve a pedagógiai intézetek által szervezett tanfolyamok sikeréhez.

a Szerkesztők

CSAPÓ BENŐ

A TANULÓI TELJESÍTMÉNYEK ÉRTÉKELÉSÉNEK
MÉRÉSES MÓDSZEREI

Lektorálta: V á r i P é t e r
Országos Pedagógiai Intézet

T A R T A L O M

BEVEZETÉS	9
1. TESZTEK AZ ISKOLÁBAN	11
1.1. Értékelés, mérés, tesztelés	11
1.2. A tesztek típusai, a pedagógiai tesztek sajátosságai	14
2. A TUDÁS TESZTELÉSE	15
2.1. A tudásszintmérés funkciói	15
2.2. A tudás tesztelésének sajátosságai	18
3. A TESZTELÉS ELMÉLETE	19
3.1. A klasszikus tesztelmélet alapjai	19
3.2. Valószínűségi tesztelméletek	24
3.3. A tesztek jószágmutatói	27
3.3.1. Objektivitás	28
3.3.2. Reliabilitás	31
3.3.3. Validitás	33
3.4. Az itemek jellemzői	37
3.5. Norma-orientált és kritérium-orientált tesztelés	38
4. FELADATÍRÁS	42
4.1. A tananyag elemzése, a felmériendő tudás kategorizálása	42
4.2. Feladattípusok	45
4.2.1. Feleletválasztó feladatok	46
4.2.1.1. Alternatív választás	47
4.2.1.2. Többszörös választás	47
4.2.1.3. Válaszok illesztése	49

4.2.2. Feleletalkotó feladatok	50
4.2.2.1. Kiegészítés	50
4.2.2.2. Rövid válasz	51
4.2.2.3. Hosszú válasz	51
4.2.2.4. Esszé típusú válasz	52
4.2.3. A képesség-jellegű tudás mérésére alkalmas feladatok	52
5. TESZTSZERKESZTÉS	57
5.1. Az értékelési rendszer kialakítása	57
5.1.1. A javítókulcs és az értékelés technikája	58
5.1.2. Pontozás, súlyozás	64
5.2. Ekvivalens tesztváltozatok készítése	62
5.3. A tesztek kipróbálása és standardizálása	64
6. TESZTELEMZÉS, TESZTFEJLESZTÉS	67
6.1. A feladatok, itemek jóságának vizsgálata	67
6.2. A reliabilitásmutatók kiszámítása	71
IRODALOM	74

B E V E Z E T É S

Kevés olyan olyan fogalom van a pedagógiában, amelyet több ellentétes vélekedés, megnyilvánulás kísérne, mint a méréses módszereket, a tesztek használatát. Az ellentmondásoknak megvannak a maguk természetes okai: nemcsak a tesztelésnek a pedagógusok körében való megítélése ellentmondásos, hanem a tesztek használatának magyarországi helyzete is. Egy-egy pedagógiai kutatóhely néhány területen a világban megfigyelhető főbb fejleményekkel egyidőben vagy azt megelőzve dolgozott ki eredeti megoldásokat, azonban még az így hozzáférhetővé vált ismeretek és technikák sem terjedtek el, a tesztek készítése és használata nem vált pedagógiai kultúránk szerves részévé.

Ugyanakkor az utóbbi évtizedekben egyre kidolgozottabbá vált a tesztelés elmélete, és a világ különböző oktatási rendszereiben rohamosan terjedtek a tudás értékelésének méréses módszerei. A gyors térhódításnak nyomós okai vannak: az oktatás hatékonyságának fokozása tovább már nem lehetséges az erőteljes differenciáció, a tanulók egyéni igényeihez és szükségleteihez jobban alkalmazkodó, személyre szabott módszerek nélkül. Ezek a módszerek pedig az elsajátítás gyakori és megbízható értékelését teszik szükségessé, ami ma már elképzelhetetlen a feladatlapok, tesztek nélkül. A fejlett oktatási rendszerrel is rendelkező vezető ipari országokban tesztkészítéssel, tesztfejlesztéssel foglalkozó hatalmas kutatóközpontok működnek és látják el az oktatást a szükséges tesztekkel, elméleti és gyakorlati ismeretekkel.

Sajnos a tesztekkel kapcsolatos ismeretek bővülésével pedagógiai könyvkiadásunk nem tartott lépést. Mintegy hatnyolc vaskos kézikönyvre lenne szükség ahhoz, hogy az ezzel foglalkozó szakemberek, specialisták számára szükséges ismeretek magyar nyelven is hozzáférhetővé váljanak. Természetesen ez nem jelenti azt, hogy minden tesztet használó pedagógusnak, pedagógiai vezetőnek is ilyen mélységű ismeretekkel kellene rendelkeznie, de azt mindenképpen érzékeltetnünk

kell, hogy sok területen az alaposabb felkészülést igénylőknek csak angol vagy német nyelvű szakirodalmat ajánlhatunk.

A mérési módszerek legalább három önálló, de szorosan összefüggő kérdéskört érintenek: az értékelés és azon belül a mérési módszerek pedagógiai elméletét, a tesztek matematikai (statisztikai, valószínűségelméleti) elméleteit, valamint a tesztek készítésének technikájával kapcsolatos ismereteket.

Az e füzetben rendelkezésre álló rövid terjedelemben természetesen nem foglalkozhatunk mindennel részletesen. Arra vállalkozhatunk csupán, hogy a tesztek készítéséről és használatáról annyit elmondjunk, amennyi megítélésünk szerint ma már minden szakmájában képzett pedagógus műveltségéhez hozzátartozik. E füzet célja az lehet, hogy áttekintést adjon, orientáljon, segítsen a kész, felhasználni kívánt tesztek megítélésében. Kialakíthatja azt az érzéket, ami segíti a jó, professzionális színvonalú és az esetleges, megbízhatatlan mérőeszközök közötti különbségtételt. Az így megszerezhető tudás segíthet abban, hogy saját céljainkra vagy szűk körű felhasználásra megfelelő feladatlapokat, tesztek készítsünk. Azonban amint ezt a később elmondandók ugyancsak bizonyítani fogják, a széles körű használatra szánt, lényeges döntéseket megalapozó tesztek készítése nagy technikai apparátust, sajátos felkészültséget igényel. Mindazoknak, akik ilyen feladatokra szeretnének felkészülni, a további speciális tanulmányokat javasoljuk.

Mondanivalónk középpontjába a tesztek készítésének (5-6.fejezet), feladatok írásának (4. fejezet) gyakorlati kérdéseit állítottuk. A legközvetlenebb gyakorlati kérdéseket (Mit mér a tesztünk? Jól méri-e azt, amit mér?) sem tudjuk azonban megválaszolni megfelelő elméleti keretek nélkül, így a 3. fejezetben a tesztelmélet alapvonásait tekintjük át. Természetesen szükség van arra is, hogy az egész vállalkozás értelmét, a tesztek funkcióit, használatuk előnyeit, szükségességét bemutassuk, így az első két fejezetben röviden ezekkel a kérdésekkel is foglalkozunk.

1. TESZTEK AZ ISKOLÁBAN

1.1. Értékelés, mérés, tesztelés

Ha a tudásszintmérés problémáit el akarjuk helyezni a pedagógiai folyamatokban, mindennek előtt az értékelés fogalmával kell megismerkednünk. Az értékelés a pedagógiai folyamatok egyik alapszakasza, a visszacsatolás funkcióját látja el. Arra hivatott, hogy informálja a pedagógiai folyamatok résztvevőit (a tanulókat, a pedagógusokat, a köznevelési rendszer irányítóit) arról, hogy a folyamat adott tevékenységei mennyire érték el céljukat. A tervszerű pedagógiai munka alapvető feltétele, hogy ilyen információkat nyerjünk, ezek nélkül nincs lehetőség arra, hogy a folyamatokat céljainknak megfelelően irányítsuk.

Szűkebb területünknel, a tanulói teljesítmények értékelésénél maradva szükség van arra, hogy időről-időre megvizsgáljuk, hol tartanak a tanulók a tananyag elsajátításában, hogyan fejlődnek készségeik, képességeik. Csak e visszajelzések birtokában tudjuk munkánkat eredményesen végezni, így tudjuk meghatározni, mi az amit a tanulók már elsajátítottak és mi az amit még tovább kell tanítani. Ugyanezek az információk a tanulók számára is nélkülözhetetlenek, nemcsak visszajelzést adnak, de a motiváció forrásai is.

Az információk felvételének, összegyűjtésének különböző módzatai vannak. Előfordul, hogy a tanulónak kérdéseket teszünk fel, és a válaszokat értékeljük, de hosszabb, összefüggő szöveg elmondására is készíthetjük őket. Problémákat tűzhetünk ki számukra, amelyeket írásban vagy szóban megoldanak; dolgozatokat, röpdolgozatokat írásthatunk, különböző alkotásaikat értékelhetjük. E sokféle értékelési módot alapve-

tően két csoportra oszthatjuk, attól függően, hogy értékelésünk pusztán minőségi természetű, vagy az értékelés eredményét valamilyen mennyiséggel számszerűen is kifejezzük. A minőségi, vagy kvalitatív értékelés nélkülözhetetlen eleme a pedagógiai folyamatoknak. Kööttségek nélkül teszi lehetővé az eredmények hosszabb, szóbeli vagy írásbeli elemzését, értékelését. A kööttségek hiánya azonban egyben azt is jelenti, hogy a minőségi értékelés során kevés a támpontunk. A véleményben sok a szubjektív mozzanat, a minőségi értékelés eredményei nehezen hasonlíthatók össze egymással.

Az értékelés másik fő formáját azok a tevékenységek jelentik, amelyeknek eredményeként a megvizsgált teljesítményekhez valamilyen számszerű értéket rendelünk hozzá, azzal fejezzük ki az adott teljesítmény színvonalát, értékét. Az értékelésnek ezt a típusát nevezzük kvantitatív, vagy mennyiségi értékelésnek. A mennyiségi értékelésnek is különböző szigorúságú fokozatai vannak. Tulajdonképpen a kvalitatív és a kvantitatív értékelés közötti átmenetnek tekinthetjük a legegyszerűbb minősítést, a megítélést. Ekkor a teljesítményeket két csoportba soroljuk, elfogadhatónak, vagy elfogadhatatlannak tartjuk. Azt mondjuk, hogy a tanuló megfelelt, vagy nem felelt meg, produktuma jó, vagy rossz. Ebben az esetben a teljesítményeket két minőségileg különböző osztályba soroljuk, de véleményünkben már a mérték képze is szerepet játszik: megítéljük, hogy az adott teljesítmény eléri-e az elfogadhatóság valamilyen többé-kevésbé rögzített mértékét. Közelebb jutunk a kvantitatív értékelés lényegéhez, ha az értékelendő teljesítményt egy skálán próbáljuk meg elhelyezni. Az ilyen módon elképzelt, vagy ténylegesen rögzített skálák még mindig sokfélék lehetnek. Ezekhez a skálákhoz viszonyíthatjuk az adott konkrét értékelendő teljesítményeket, a skála fokai-val fejezzük ki a teljesítmény színvonalát.

Ha nincs módunk arra, hogy egy ilyen skálát hozzámérjünk az adott teljesítményhez, hanem csak az általunk elképzelt skálán gondolatban helyezzük el a megfigyelt teljesítményt,

akkor becslést végzünk. A becslés mindennapi életünk egyik leggyakoribb tevékenysége. Igen gyakran nincs szükség arra, hogy mérőeszközzel pontosan meghatározzunk egy adott mennyiséget, például egy hosszúságot, elegendő, ha a hosszúság egységének, a méternek a képzetét gondolatban mérjük hozzá az adott tárgyhöz. Így becsléssel meghatározhatjuk például egy fának a magasságát. Hasonlóképpen becslést végzünk akkor is, amikor a tanulókat osztályozzuk. Az iskolában szokásos öt fokozatú skála minden pedagógus képzeletében egy mértékrendszernek felel meg, a pedagógusok kialakítják az egyeshez, ketteshez, hármashoz, négyeshez, ötöshöz tartozó értékeket, és képzeletben az adott tanuló feleletét, dolgozatát ezen a skálán helyezik el. Nyilvánvaló, hogy ebben az esetben sem az "ötösség", sem a "négyesség" mértéke nem szigorúan rögzített, de már az ily módon elvégzett becslés is eleget tesz bizonyos, a mérésekkel szemben is támasztható követelményeknek. Például, ha ezeket a becsléseket jól végezzük, akkor nem fordulhat elő az, hogy két megfigyelt teljesítmény közül ahhoz, amelyik jobb volt, kisebb osztályzat tartozzon. Ezt a követelményt még egy viszonylag lazán körülhatárolt követelményrendszer esetén is teljesíteni lehet.

A megvizsgálandó mennyiségekről sokkal pontosabb adatokat kaphatunk mérés segítségével. A mérést úgy végezzük, hogy egy mérőeszközzön rögzített skálát hozzámérünk egy megvizsgálandó tulajdonsághoz.

Ha pedagógiai mérésekről beszélünk, ott is a mérés hétköznapi fogalmából indulhatunk ki. Elkészítünk egy mérőeszközt, a mérőeszközzön létrehozuk a megfelelő skálát, majd ezt a mérőeszközt illesztjük hozzá a megvizsgált tulajdonsághoz, teljesítményhez. A pedagógia mérések tipikus eszköze a teszt. A teszt szó eredeti értelmében próbát jelent; a pedagógiai és pszichológiai szóhasználatban a teszt, a tesztelés a megvizsgálandó tulajdonság kipróbálása útján való értékelését jelenti. A következőkben minden olyan pedagógiai mérőeszközt, amely eleget tesz a vele szemben támasztható követelmények-

nek, - ezekről később lesz szó -, tesztnek fogunk nevezni.

A tesztek használata elsőként a pszichológiában vált népszerűvé. Segítségükkel lehetőség nyílt bizonyos pszichikus tulajdonságok, például az intelligencia, a kreativitás, vagy bizonyos jellemvonások gyors és többé kevésbé megbízható értékelésére. A pszichológiai tesztek közel egy évszázados múltra tekinthetnek vissza.

1.2. A tesztek típusai, a pedagógiai tesztek sajátosságai

A teszteket sokféle szempont alapján csoportosíthatjuk. Ha a pszichológiai és a pedagógiai teszteket akarjuk egymástól megkülönböztetni, nem is annyira tulajdonságaikban, formai sajátosságaikban kell a különbségeket keresni, hanem a használatuk módjában és céljában. Amíg a pszichológiai használatának célja főleg az, hogy bizonyos állapotokat, tulajdonságokat személyiségvonásokat regisztráljanak, és ennek segítségével az emberek bizonyos tevékenységre való alkalmasságára előrejelzéseket tegyenek, személyiségvonásokat meghatározzanak, leírjanak, tulajdonságok alapján jelöltek közül válogassanak, addig a pedagógiai tesztek alapvető célja az, hogy az oktatási, nevelési folyamatok irányításához információt szolgáltatassanak, és így a fejlődést, a fejlesztést, a változtatást segítsék, az ilyen irányú tevékenységet megalapozzák. Az iskolában használatos tesztekkel a tanulók sokféle tulajdonságát, személyiségvonását értékelhetjük. A szűkebb értelemben vett személyiségtesztekkel a tanulók jellemét, például introvertáltságának mértékét, moti-

váltságát, szorongásait és még sok más személyiségvonását értékelhetjük megbízhatóan.

A tesztek másik nagy csoportját képezik azok, amelyekkel a tanulók tudását értékeljük. Ezek a tesztek a szűkebb értelemben vett oktatásban játszanak szerepet. Segítségükkel a tananyag elsajátításának színvonalát, képességek, készségek fejlődését értékelhetjük. A későbbiekben csak ezekkel a tesztekkel, a tudás mérésére alkalmas tesztekkel foglalkozunk.

2. A TUDÁS TESZTELÉSE

2.1. A tudásszintmérés funkciói

Mielőtt a tesztelés elméleti kérdéseivel foglalkoznánk, választ kell adnunk arra az alapvető kérdésre: miért is van szükség az iskolákban a tesztekre? Mindenki óhatatlanul felmerül a kérdés, vajon valóban szükség van-e arra, hogy a tanulók tudását személytelen eszközökkel vizsgáljuk. Vajon miért nem elegendő az értékelés hagyományos eszközszerete, a tanár megítélése, a felelés, a dolgozat, a röpdolgozat? Ha a tesztelés szükségességét akarjuk jellemezni, rohamos terjedésük okait akarjuk megérteni, szükség van arra, hogy a tesztelés kérdéseit az oktatási rendszerek tágabb összefüggéseiben vizsgáljuk.

A tesztek az oktatásban, az oktatási folyamat különböző szakaszaiban és szintjein jelennek meg. Elsőként a tanítástanulás mindennapos folyamatait vizsgáljuk meg. A tanítástanulás ősi modellje az, amikor egy pedagógus egy tanulóval

áll kapcsolatban. Ebben az esetben a pedagógus képes arra, hogy egyetlen tanulójának tanulását pontosan irányítsa, azaz megfigyelje, hogy mi az, amit a tanuló már elsajátított, és mi az, amit még tanulnia kell; mi az, ami még további magyarázatot igényel. Az oktatás tömegessé válásával azonban ez az ősi modell felbomlott. Egy pedagógus a tanulók 20-30 fős csoportjával áll kapcsolatban, és ezek között a tanulók között óriási különbségek lehetnek. Ilyen körülmények között a pedagógus nem képes arra, hogy minden egyes tanítványának a tudását kellő részletességgel és kellő gyakorisággal megvizsgálja. Hagyományos módszereivel csak a tanulók tudásának egy töredékét képes feltérképezni. Egy osztálynyi gyereket már csak úgy tud egyszerre vizsgáztatni, ha egyszerre mindannyiuknak írásbeli munkát ad, és munkájuk eredményét utólag értékeli. Ebből a szemszögből nézve a teszt nem más, mint egy sajátos dolgozat, amely nagyon célszerűen megválogatott feladatokat tartalmaz, a feladatok eredményeit pedig gyorsan, egyszerűen, egyértelműen, megbízhatóan lehet értékelni. Ebben a szerepükben a tesztek a tanítási-tanulási folyamatok megbízhatóbb irányítását segítik. Talán nem véletlen, hogy mindazok az oktatási modellek, melyek az oktatást hatékonyabbá akarják tenni, a tananyag teljes, magasszintű elsajátítását célozzák (mastery learning, perszonalizált oktatás), azok a tesztek megfelelő gyakoriságú használata nélkül nem is működnek.

Természetesen a teszteknek (miként általában a pedagógiai értékelésnek), vannak más funkciói is. Bizonyos esetekben (például az iskolafokokozatok közötti átmenet esetében) szükség van arra, hogy nagy tömegű tanuló tudását viszonylag rövid idő alatt, gyorsan, megbízhatóan mérjük fel. Ilyen esetben a tanulók szelektálását, kategorizálását segíthetik a tesztek. Magyarországon egyetlen ilyen tesztforma működik - az egyetemi, illetve főiskolai felvételi teszteknek van ilyen funkciójuk. Tesztek nélkül nem működhetne megbízhatóan a közoktatási rendszer irányítása

sem, tesztek nélkül nem lehetne a kutatókat, döntéshozókat megfelelő információkkal ellátni. A tesztek használata által viszont - mivel azok a tudást egy adott egységes skálán mérik - lehetővé válik, hogy térben és időben egymástól távol eső teljesítményeket mérjünk össze. Egy Zala megyei osztály tanulóinak tudását ugyanazon a skálán fejezhetjük ki, mint Békés megyei társaikét, és így tudásukat összehasonlíthatjuk. Ha a tesztünk jó, ez az összehasonlítás nem függ tanáraik egyéni véleményétől, illetve attól, hogy milyen szigorúsággal osztályoznak, milyen saját képzetet alakítottak ki a tanulók teljesítményéről, hova helyezik a négyes és az ötös közötti határt.

Tesztek segítségével megbízhatóan összehasonlíthatjuk különböző oktatási eljárások hatékonyságát, de segítségükkel értékelhetjük, össze-mérhetjük a különböző tantervek időben egymástól távoli eredményeit is. Ma például már dokumentált adataink vannak arról, hogy 20 évvel ezelőtt milyen volt az általános iskolai tanulók számolási készsége. Ha a mai tanulókat ugyanazokkal a tesztekkel felmérjük, mint amelyekkel a 20 évvel ezelőtti méréseket végezték, lehetőségünk van például annak a meddő vitának a megbízható lezárására is, hogy vajon ma tényleg gyengébben számolnak-e általános iskolásaink, mint régebben. Megfelelő tesztek nélkül nem tudnánk összehasonlítani különböző országok oktatási rendszereinek eredményességét sem.

2.2. A tudás tesztelésének sajátosságai

Mivel az első tesztek a pszichológiában terjedtek el, természetes, hogy először a pedagógiai tesztek, közelebbről a tudásszint mérésére alkalmas tesztek is a pszichológiai tesztek formai sajátosságait másolták le. Később aztán egyre inkább előtérbe kerültek azok a tulajdonságok, amelyekkel a tudásszintmérő teszteknek rendelkezniük kell, és ma már e teszteknek önálló elméletük van. Célszerű röviden áttekintenünk, melyek azok a sajátosságok, amelyekkel a tudásszintmérő tesztek rendelkeznek. A tudásszintmérő tesztek szinte kivétel nélkül a m a x i m á l i s t e l j e s í t m é n y t m é r i k , azaz azt vizsgálják, mire képes a tanuló, ha minden tudását összeszedi, képessége legjavát nyújtja. Ez sok pszichológiai tesztnél nem így van. Néhány pszichológiai teszt nem a maximális teljesítményt, hanem a tipikus teljesítményt vizsgálja, és a tulajdonságot, az adott személy tulajdonságát két pólus között próbálja meg valahol elhelyezni. Ilyen tulajdonságpárok például az introvertált-extrovertált, vagy a maszkulin-feminin jelleg.

A tudásszintmérő-tesztek sajátossága az is, hogy többségük használata c s a k a z o k t a t á s i f o l y a m a t t a l s z e r v e s e g y s é g b e n képzelhető el. Pontosan meg van határozva, melyik az az időpont, amikor az adott teszt használatának értelme van, míg pszichológiai vizsgálatokra a legkülönbözőbb alkalmakkor kerülhet sor. Ez a körülmény a tudásszintmérő-tesztekkel kapcsolatban sajátos kötöttségeket jelent.

Tudásszintmérés esetén többnyire pontosan körül tudjuk határolni azt a tudást, amit száz százaléknak tekintünk, és amihez az egyes tanulók tudását viszonyítjuk. Sok személyiségvonás esetében nem tudjuk megmondani, mi a teljesség, a v i s z o n y í t á s i p o n t .

Fontos különbség az is, hogy a pszichológiai teszteknel

általában minden egyes kérdés, feladatelem ugyanazt a tulajdonságot méri, míg a tudásszintmérő tesztek a tudás egy többé - kevésbé összefüggő területének különböző elemeit.

3. A TESZTELÉS ELMÉLETE

3.1 A klasszikus tesztelmélet alapjai

A tesztelméletek matematikai, elsősorban valószínűségi és statisztikai összefüggések rendszerei. A tesztelméletek gyökerei e század első évtizedeire nyúlnak vissza, teljesen kifejlett formájukban pedig az 50-es évek végén jelentek meg. A tesztelméletek alapfeladatuknak tekintették, hogy a tesztekkel kapcsolatos matematikai illetve statisztikai törvényszerűségeket egységes rendszerbe foglalják. Néhány alapvető kiinduló feltevést rögzítenek, majd ebből a szokásos matematikai eljárásokkal tételeket vezetnek le, amely tételeket azután a tesztekkel kapcsolatos gyakorlati kérdések megválaszolásában is használhatunk.

Ma már klasszikus tesztelméletként szokás említeni - az újabb tesztelméletektől való megkülönböztetésként - azt a tesztelméletet, amelynek alapgondolatait itt felvázoljuk. Ez az elmélet is néhány alapfogalomra és néhány alapfeltevésre épül, magát a tesztelméletet a belőlük levezetett összefüggések rendszere alkotja.

Az alapfogalmak közül természetesen először a teszt fogalmát kell tisztáznunk. A teszt ebben az értelemben pszichológiai (pedagógiai) mérőeszközt jelent, amely a pszichikus

tulajdonságokat megfelelő skálán méri. A teszt kisebb, önállóan értékelhető egységekből áll. Ez lehet egy szubteszt (rész-teszt), amely a teszt kisebb, önálló feladatokból álló részét jelenti. (Nem feltétlenül szükséges, hogy a tesztet szubtesztekre bontsuk, másrészt a szubteszteket tekinthetjük önálló teszteknek is.) A tesztek feladatokból állnak, a feladatok legkisebb, önállóan értékelhető részeit itemeknek nevezzük. A pedagógiai, ezen belül a tudásszintmérő tesztekben hasznos ennek a négy szintnek a megkülönböztetése. A legtöbb pszichológiai teszt azonban nem bomlik fel ilyen sok szintre, a klasszikus tesztelmélet feltevése szerint elegendő csak a tesztek és az azokat alkotó legkisebb önálló egységeket, az itemeket értelmezni.

Az itemek, mivel a legkisebb önállóan értékelhető egységek, már nem bonthatók fel további részteljesítményekre, éppen ezért az itemeket már csak kétféleképpen minősíthetjük: a tanuló megoldotta, az item jó, a tanuló nem oldotta meg, az item nem jó. A klasszikus tesztelmélet szerint 1 pontot ér az item, amelyet a tanuló jól oldott meg, 0 pontot természetesen az, melyet a tanuló nem jól oldott meg. (Tudásszintmérő tesztek esetében az itemeket súlyozhatjuk is, azaz nem csak 1 pontos, hanem az item súlyától függően több pontos itemeket is definiálhatunk, ld. később.) A tesztek összpontszámát, vagyis a teszt eredményét az itemekre kapott pontok összege adja. Ha tehát 0-val és 1-gyel pontozunk, a teszt összpontszáma a jól megoldott itemek számával lesz egyenlő.

A tesztelmélet célja az, hogy a tesztekkel kapcsolatban felmerülő kérdéseket matematikai úton fogalmazza meg, és ezáltal lehetővé tegye, hogy kérdéseinkre egzakt választ kapjunk. Az egyik legtermészetesebben felmerülő gyakorlati kérdés az, hogy tesztünk mennyire jól méri azt, aminek a mérésére kidolgoztuk. Példaként nézzük meg, hogy erre a kérdésre milyen módon keresi a választ a klasszikus tesztelmélet. Ehhez mindenekelőtt fel kell írunk a klasszikus tesztelmélet alapegyenletét, amit a következő formában tehetünk meg. Fel

kell tételeznünk, hogy annak a tulajdonságnak, amelyet mérni szeretnénk, minden személyre jellemzően van egy meghatározott valódi értéke. Tehát minden vizsgálandó személyhez hozzátartozik például valamilyen intelligenciaszint, a kreativitásnak valamilyen mértéke, vagy témánk fogalmait használva egy adott tárgy valamilyen tudásszintje. Ezt nevezzük az adott tulajdonság, - például a tudásszint - valódi értékének, amit V betűvel jelölünk. Fel kell azonban azt is tételeznünk, hogy amikor ezt a tulajdonságot mi a tesztünkkel meg akarjuk mérni, akkor a tesztrel való mérés során, - mint ahogy mindenféle mérés során - hibát követünk el. Tesztünk tehát nem pontosan tükrözi ennek a tulajdonságnak a valódi értékét. Ezt a klasszikus tesztelmélet úgy veszi figyelembe, hogy egy bizonyos mértékű hibát definiál. Ezt a hibát H betűvel jelöljük. Ily módon, amikor egy tulajdonságot megmérünk, az általunk mért értékben a valódi érték, és a hiba játszik szerepet. Ez a hiba a valódi értéket felfele vagy lefele eltérítheti, felfele például úgy, hogy a tanuló véletlenszerűen ráhibáz a jó megoldásra, lefele úgy, hogy figyelmetlen, nem a tudása legjavát nyújtja. Ezt az alapvető feltételezést a klasszikus tesztelmélet alapegyenletében foglalhatjuk össze, következőképpen:

$$M=V+H$$

A képletben M a mért értéket, a V a tulajdonság valódi értékét, H pedig a hibát jelöli. Ehhez az egyenlethez a klasszikus tesztelmélet bizonyos feltételezéseket kapcsol. Olyan feltevéseket, amelyeket igaznak fogadunk el, és amely feltevések igazsága esetén az összes azokból levezetett összefüggések is mind igazak. Itt nem soroljuk fel tételesen a klasszikus tesztelmélet axiómáit, csupán utalunk arra, hogy ezek milyen jellegűek. Feltesszük például azt, hogy a hiba várható értéke nulla, vagyis a valódi értéktől ugyanolyan valószínűséggel térnek el a tanulók vagy a megvizsgált személyek negatív, mint pozitív irányban, tehát összességében a sokféle hiba statisztikusan kiegyenlíti egymást. Feltesszük

továbbá azt is, hogy a méréskor jelentkező hiba nem függ attól, hogy milyen az adott tulajdonság valódi értéke. Statisztikailag ezt úgy fogalmazhatjuk meg, hogy a valódi érték és a hiba közötti korreláció 0. Az ehhez hasonló további alapfeltevések rendszeréből a matematikai levezetések szabályait figyelembe véve a tesztelmélet tételeit vezethetjük le. Visszatérve kiinduló kérdésünkhöz - tehát hogy mennyire jól méri a teszt azt, amit vele mérni kívánunk - e kérdést a klasszikus tesztelméletet felhasználva matematikai formában fogalmazhatjuk meg. Azt mondhatjuk, hogy tesztünk annál jobban méri a mérni kívánt tulajdonságot, minél jobban összefügg a valódi érték a mért értékkel. Statisztikai kifejezéssel: minél jobban korrelál M és V .

A tesztnek ezt a tulajdonságát, az így értelmezett jószágát a teszt megbízhatóságának, idegen szóval reliabilitásának nevezzük. A reliabilitás tehát a klasszikus tesztelmélet definíciója értelmében a mért érték és a valódi érték közötti összefüggés. Pontos matematikai meghatározása szerint a reliabilitási koefficiens amit r_{tt} -vel jelölünk, a mért és a valódi érték közötti korrelációs együttható négyzete:

$$r_{tt} = r^2.$$

Szokás még az úgynevezett reliabilitási indexről beszélni, ami nem más, mint az előbbieken definiált korrelációs együttható.

Kérdésünket ezzel matematikai pontossággal fogalmaztuk meg, gyakorlati problémánk azonban ezzel még nincs megoldva. Ha pontosan megfigyeltük a gondolatmenetet, abból kitűnik, hogy egy olyan korrelációs együtthatót kellene kiszámítanunk, aminek az értékét soha nem tudjuk meghatározni. Nem tudjuk ugyanis meghatározni a korrelációs együtthatóban szereplő egyik mennyiséget, vagyis az adott tulajdonság valódi értékét. Bármilyen tesztet használunk, minden tesztben a valódi érték mellett valamilyen hibát is mérünk, tehát mindig csak bizonyos mért értékekkel számolhatunk. Ezen a ponton derül ki, hogy mi az értelme az egész formalizált tesztelméletnek, a

matematikai apparátus felhasználásának. Lehetőségünk van ugyanis arra, hogy ismert matematikai, statisztikai tételek felhasználásával olyan formulákat dolgozzunk ki - a már előbb említett axiómákra alapozva -, amelyekben már csupa mért értékek szerepelnek, és amely formulákkal becslést adhatunk a reliabilitásra.

Mivel ez az egyik legfontosabb gyakorlati probléma (azaz, hogy tesztjeink megbízhatóságát jellemezzük), a klasszikus tesztelmélet keretében számos reliabilitási mutatót dolgoztak ki. Ezek a reliabilitási mutatók olyan formulák, amely formulákba csupa mért értéket kell behelyettesíteni, és segítségével a reliabilitási értéket lehet kiszámítani. Pontosabban a reliabilitásra lehet egy alsó becslést adni. Olyan becslést, amelynél a reliabilitás valódi értéke mindig csak nagyobb lehet. Ezeket a becsléseket a megfelelő statisztikai törvényszerűségekre alapozhatjuk. A reliabilitás fogalmával, gyakorlati értelmezésével és a formulák felhasználásával a következő fejezetben még részletesebben foglalkozunk. Itt csupán illusztrálni szerettük volna a klasszikus tesztelmélet matematikai apparátusának működését. További problémákat a klasszikus tesztelmélet hasonló formulákkal, levezetésekkel old meg.

A 60-as évek végére, mikorra a klasszikus tesztelmélet végleges formáját elérte, az is kiderült, hogy a gyakorlatban felmerülő számos problémára nem tud választ adni, illetve, hogy azok a feltételek, amelyeket az axiómákban rögzített, a pszichológiai tesztek használata során többnyire nem teljesíthetők, vagy nem szigorúan teljesíthetők.

Még több probléma jelentkezett a pedagógiai tesztek esetében, különösen a tudásszintméréssel kapcsolatban. Itt az a helyzet, hogy azok az értékek, amelyeket a pszichológiai tesztek értéknek tekintenek, nem feltétlenül értékek a pedagógiai, a tudásszintmérő tesztek használatánál. Csak egy példát említve: a pszichológiai tesztek akkor működnek jól, ha azokon a megvizsgált személyek közel normális eloszlású tel-

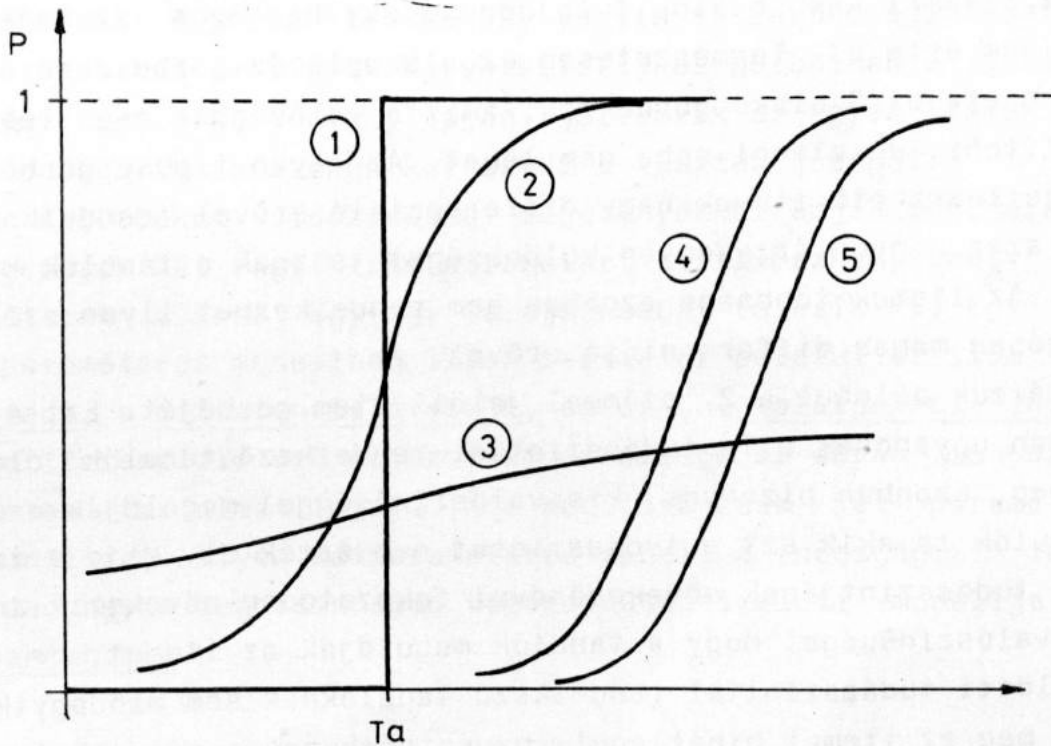
jesítményt produkálnak. A tesztek különböző technikai megoldásokkal, úgy készítik el, hogy a végső eredményeik már normális eloszlást adjanak. Ezzel szemben egy pedagógiai tesztnél (pl. egy tudásszintmérő tesztnél) nem feltétlenül kell elérnünk azt, hogy normális eloszlásúak legyenek az eredmények, hiszen elképzelhető az is, hogy a tanulók nagy többsége, vagy szinte mindenki tökéletesen megoldja a tesztet. Ha a tanítás során jól dolgoztunk, akkor éppen erre törekszünk. Szélsőséges esetben, ha mindenki 100%-os, tökéletes megoldást nyújt, akkor a tesztnek nem is lesz szórása, tehát a klasszikus tesztelmélet összes formulája használhatatlanná válik. Ezek a problémák vezettek el oda, hogy különböző, más megfontolásokon alapuló tesztelméleteket dolgozzanak ki. A tesztelmélet modern irányzatai az úgynevezett probabilisztikus (valószínűségelméleti) tesztelméletek. Ezekkel röviden a következő pontban foglalkozunk.

A tesztelmélet fejlődésének a másik iránya már szorosabban kapcsolódik a tudásszintméréshez, ez pedig a kritérium-orientált tesztelmélet, amivel ugyancsak később foglalkozunk részletesen.

3.2. Valószínűségi tesztelméletek

A tesztelméletek újabb generációját képezik a modern, ún. probabilisztikus (valószínűségi) tesztelméletek. Ezek a tesztelméletek a teszt használata során elkövethető hibát más módon veszik számba, mint tette azt a klasszikus tesztelmélet. Mivel ezeknek a tesztelméleteknek meglehetősen bonyolult a matematikai apparátusuk, csupán az alapgondolatukat mutatjuk be. Az elmélet a tesztitemek tulajdonságait valószínűségelméleti eszközökkel jellemzi. Ebben a tesztelméletben az, hogy egy tanuló megold egy tesztitemet, vagy nem old meg, nem

determinisztikus, hanem valószínűségi természetű. Ez azt jelenti, hogy ha különböző tudásszintű tanulókat veszünk, akkor annál nagyobb annak a valószínűsége, hogy az adott itemet a tanuló jól fogja megoldani, minél nagyobb a tanulónak a tudásszintje. Az összefüggés azonban nem determinisztikus, hanem valószínűségi, azaz a gyengébb tudású tanuló kisebb valószínűséggel, a jobb tudású tanuló nagyobb valószínűséggel oldja meg az adott tesztitemet. Így az item tulajdonságait jól kifejezi az itemhez rendelhető valószínűségeloszlás. A valószínűségeloszlás és az itemek tulajdonságai közötti összefüggéseket az 1. ábrán szemléltetjük.



1. ábra

Különböző paraméterekkel rendelkező itemek
karakterisztikus görbéi

Az 1. ábrán öt különböző item valószínűségi függvényét ábrázoltuk. Az x tengelyen a tanulók tudását tüntettük fel,

az y tengelyen pedig annak valószínűségét, hogy a tanuló az itemet jól oldja meg. A valószínűségeloszlást ebben az esetben jellegzetes S alakú görbék tükrözik. A görbék alakja jól kifejezi azt, hogy milyen tulajdonságú itemmel állunk szemben.

Az 1. számmal jelölt görbe például egy olyan itemnek a karakterisztikáját jelöli, amelyet senki nem tud megoldani, aki nem érte el legalább a T_a tudásszintet, de mindenki meg tud oldani, aki ennél a T_a -nál nagyobb tudásszinttel rendelkezik. Ez tehát egy olyan item, ami nagyon pontosan két csoportra osztja a tanulókat aszerint, hogy a tanuló elérte-e az adott itemmel kapcsolatos tulajdonság egy bizonyos szintjét, vagy nem érte el. Természetesen ez a szögletes görbe csak egy határeset az S alakú görbéknek, amit a valóságban meg lehet közelíteni, de elérni soha nem lehet. Az ilyen típusú görbékkel jellemezhető itemek nagy differenciáló erővel rendelkeznek, azaz nagyon érzékenyen különbséget tesznek a tanulók között. Az itemek többsége azonban nem rendelkezhet ilyen szélsőségesen magas differenciáló erővel.

Nézzük például a 2. számmal jelölt item görbáját. Ezt átlagosan ugyancsak a T_a tudásszinttel rendelkező tanulók oldják meg, azonban bizonyos, kis valószínűséggel megoldják azok a tanulók is akik ezt a tudásszintet nem érték el. Majd a tanulók tudásszintjének növekedésével fokozatosan növekszik annak valószínűsége, hogy a tanulók megoldják az itemet, de a T_a fölötti tudásszinttel rendelkező tanulóknak sem mindegyike oldja meg az itemet hibátlanul, hanem csak egyre nagyobb valószínűséggel. Azonban még az igen magas tudásszinttel rendelkező tanulókról sem mondhatjuk biztosan, hogy megoldják az itemet, hanem csak azt, hogy ezek esetében a megoldás már nagyon valószínű. Ez az item láthatóan kisebb differenciáló erővel rendelkezik, mint az 1. item.

Ha megnézzük a 3. számú item görbáját, azt látjuk, hogy ezt az itemet csaknem ugyanolyan valószínűséggel oldják meg a gyenge tanulók, mint a magas tudásszinttel rendelkező tanu-

lók. Ez egy lapos karakterisztikus görbe, ami azt jelenti, hogy ez az item nem túlságosan differenciál a tanulók között. Ha tehát olyan tesztet készítünk, amelynek az a lényege, hogy a tanulókat tudásuk szerint jól megkülönböztesse egymástól, akkor az ilyen jellegű itemeket ki kell hagyni a tesztből. Természetesen más a helyzet, ha a célunk alapvetően nem az, hogy a tanulók között igen érzékenyen különbséget tegyünk. A 4. és az 5. itemek összehasonlítása két párhuzamosan futó karakterisztikus görbét mutat. Az 5. számú görbe olyan, mintha a 4-et a tanulók tudásszintjét jellemző tengely mentén a pozitív irányba eltoltuk volna. Ez azt jelenti, hogy a két item nehézségében van csak különbség, többi jellemzőjük egyébként megegyezik. A probabilisztikus, vagyis valószínűségi tesztelméletek tehát az itemek karakterisztikus görbéinek a jellemzőit vizsgálják, annak alapján jellemzik az egyes itemek tulajdonságait, és ezeken keresztül a tesztek jóságát.

A görbéket valószínűségi függvényekkel adják meg, ezek a függvények két vagy három paramétert tartalmaznak, amely paraméterek a teszt egy-egy tulajdonságát fejezik ki. Egy háromparaméteres modellben ilyen paraméter például az item nehézsége, a differenciáló ereje, és az ún. találgatási index. A differenciáló erő és az item nehézsége az ábrán is szemléletes tartalmat nyert. Más modellek csak két paramétert használnak, ilyen kétparaméteres modell a pedagógiai tesztelés egyik leghasznosabb valószínűségelméleti modellje, a Rasch modell.

3.3. A tesztek jóságmutatói

A tesztek minőségét, jóságát különböző jóságmutatókkal jellemezhetjük. A tesztek hagyományosan megkülönböztetett három jóságmutatója az objektivitás, a validitás és a reliabilitás.

Másként fogalmazva: a tesztnek objektívnek, azaz tárgyyszerűnek kell lennie; megfelelő validitással kell rendelkeznie, azaz érvényesnek kell lennie; és megfelelő reliabilitással kell rendelkeznie, vagyis megbízhatónak kell lennie ahhoz, hogy céljainknak megfelelően használhassuk. Vegyük sorra ezeket a tulajdonságokat!

3.3.1. Objektivitás

A teszt objektivitása - mint általában az objektivitás - azt jelenti, hogy a teszt tárgyyszerű, tárgyilagos, nem szubjektív. Pontosabban fogalmazva: a teszttel végzett mérés eredménye független attól, ki végzi a teszttel a mérést. Bármelyki használja a tesztet egy másik személy vagy egy csoport vizsgálatára, mindenki azonos eredményt kell, hogy a teszt segítségével kapjon. Hangsúlyozzuk, hogy itt a teszt felhasználójáról van szó, tehát nem arról a személyről, akit a teszttel mérünk, akit tesztelünk. Másként fogalmazva tehát ez azt is jelenti, hogy az elért eredményeket, a t e s z t p o n t o k a t c s a k i s a v i z s g á l t s z e m é l y t u l a j d o n s á g a i h a t á r o z z á k m e g , é s f ü g g e t l e n e k a t t ó l , h o g y a t e s z t e t k i h a s z n á l j a , k i é r t é k e l i , k i v é g z i e l a v e l e k a p c s o l a t o s s z á m í t á s o k a t .

Az objektivitásnak további formáit különböztetjük meg, attól függően, hogy a tesztelés különböző fázisaiban milyen személyek kerülnek a teszttel kapcsolatba. Ennek megfelelően az objektivitás három alapformája (alesete) az adatfelvételi objektivitás, a kiértékelési objektivitás és az interpretációs objektivitás.

Az előző definíció értelmében a z a d a t f e l v é t e l i o b j e k t i v i t á s a z t j e l e n t i , h o g y a t e s z t e r e d m é n y f ü g g e t l e n a z

a d a t f e l v e v ő s z e m é l y é t ő l . Adatfelvétele-
len azt a folyamatot értjük, amikor a teszttel az adatokat
összegyűjtjük, tehát amikor a tesztet a vizsgált tanulók meg-
oldják. Egyszerű hétköznapi példára lefordítva: bármelyik pe-
dagógus oldatja meg a tanulókkal a tesztet, ugyanazok a tanu-
lók minden pedagógus felügyelete, irányítása alatt ugyanolyan
eredményt érnek el. A teszt eredménye tehát nem függhet at-
tól, hogy ki az a pedagógus, aki a teszt megoldatásában köz-
reműködik. Első pillantásra úgy tűnik, hogy ezt a feltételt
nagyon egyszerű teljesíteni, hiszen nem kell mást tenni, mint
egyértelműen előírni a tesztet megoldató pedagógus feladatát.
Ha az utasításoknak pontosan eleget tesz, akkor a teszt meg-
oldása máris független attól, hogy ki az, aki ezeket az uta-
sításokat pontosan követi. Tudjuk azonban, hogy a gyakorlat-
ban néha nehéz ezt megvalósítani. Egy teszt megoldása során a
tanulók a legkülönbözőbb kérdésekkel fordulhatnak a pedagó-
gushoz, és így a különböző pedagógusok beállítódásuktól, fel-
adatuk értelmezésétől függően más és más válaszokat adhatnak
a tanulóknak. Egy-egy mondat, egy figyelmeztetés, egy orien-
táló megjegyzés egy egész osztály teljesítményét tolhatja el
valamilyen irányba.

Különösen kiélezetté válik az adatfelvételi objektivitás
kérdése, ha a tesztet megoldató pedagógus valamilyen módon
maga is érintett a teszt eredményében. Például úgy érzi, hogy
a teszttel egyben az ő munkáját is értékelik. Ebben az eset-
ben segíthet abban, hogy a tanulók jobb eredményeket érjenek
el, azaz a tesztek eredménye nem lesz objektív.

A teszt objektivitását ilyen helyzetben csak azáltal
biztosíthatjuk, ha megnyerjük a pedagógus együttműködő kész-
ségét és emellett feladatát pontosan körülírjuk. Pontosán
rögzítjük azt, hogy milyen utasításokat adhat a tanulóknak,
mely kérdéseikre válaszolhat, és pontosan meghatározzuk, hogy
ezen túl semmi mást nem szabad a tanulókkal közölnie. Gyakran
a pedagógus maga is abban érdekelt, hogy hiteles képet kapjon
tanítványai tudásáról, hiszen neki is fontos lehet az, hogy

megtudja, osztályának tudása hogyan viszonyul mondjuk az országos átlaghoz. Olyan esetben, amikor a teszt egyben magáról az iskoláról, az iskolai tanulók tudásáról, a pedagógus munkájáról is képet ad, vagy a pedagógus úgy érzi, a teszt egyben az ő munkájának színvonalát is méri, helyesebb, ha a teszt megíratására független szakértőket kérünk fel.

A teszt kiértékelése objektív, azaz rendelkezik a kiértékelési objektivitás tulajdonságával, ha eredménye független attól, hogy a tesztek javítását, kódolását, azaz értékelését ki végzi. Ez csak akkor lehetséges, ha a teszt értékelése teljesen egyértelmű, olyan szigorúan szabályozott, hogy ezeket a szabályokat követve bárki ugyanarra az eredményre jut. A tesztek kiértékelési objektivitását különböző eljárásokkal fokozhatjuk. A legjobb kiértékelési objektivitást akkor kapjuk, ha feleletválasztós tesztet alkalmazunk, hiszen ekkor a tanuló a feleletek közül egyszerűen csak választ, a javítókulcsok alapján pedig bárki ellenőrizni tudja, hogy a tanuló a megfelelő, azaz a helyes választ jelölte-e meg. Ilyenkor a kiértékelés objektivitását csak a kiértékelő figyelmetlensége csorbíthatja. Sajnos éppen ezzel a feladat-típussal kapcsolatban (mint később látni fogjuk) más problémák merülhetnek fel, ezért biztosítanunk kell a feleletalkotó kérdéstípusok kiértékelési objektivitását is. A feleletalkotó kérdéseknél a javítókulcs precíz megadásával, a helyes válasz pontos, egyértelmű rögzítésével biztosíthatjuk a kiértékelés objektivitását.

A teszt kiértékelési objektivitását viszonylag könnyen ellenőrizhetjük, azaz tapasztalati úton is megvizsgálhatjuk, mennyire objektív az általunk alkalmazott teszt és a hozzá tartozó értékelési utasítás (azaz javítókulcs). Nem kell mást tennünk, mint ugyanazokat a tesztekét különböző pedagógusokkal értékeltetni. A pedagógusok az értékelést a tesztől független külön lapon végzik el, az eredményeket azután szá-

mítógépbe visszük. Ugyanazokat a tesztek két, esetleg három vagy négy különböző személy is értékeli. Ha a teszt értékelése objektív, akkor minden egyes pedagógusnak, minden egyes értékelőnek azonos eredményre kell jutnia. Ha olyan kérdést találunk, amelynél az értékelők bizonytalankodnak, azaz nem mindegyikük azonosan minősíti a megoldást, alapos gyanúnk lehet arra, hogy a javítási utasítással valami probléma van, nem egyértelműen írja körül a helyes megoldást, tehát a feladat javítókulcsán (vagy a feladaton is) változtatnunk kell.

Az objektivitás harmadik formája az interpretációs objektivitás. Az interpretáció azt jelenti, hogy a tesztek eredményeit hogyan hasznosítjuk, abból milyen következtetéseket vonhatunk le, hogyan értelmezzük a teszt eredményét. Az interpretáció kérdése többnyire csak a bonyolultabb személyiségtesztek, vagy a tudományos kutatásban használatos tesztek esetében merül fel, így ezekkel a kérdésekkel itt részletesebben nem foglalkozunk.

3.3.2. Reliabilitás

A teszt egyik legfontosabb tulajdonsága a megbízhatóság, idegen szóval reliabilitás. A teszt megbízhatóságának mérésére, számszerű jellemzésére az ún. reliabilitási mutatók szolgálnak. A reliabilitási mutatók elvi kérdéseivel, matematikai statisztikai értelmezésükkel az előzőekben már foglalkoztunk. Mind a klasszikus tesztelmélet, mind pedig a modern probabilitisztikus tesztelméletek értelmezik a reliabilitást. Később a reliabilitási mutatók kiszámításának technikai részletei is sorra kerülnek, egyelőre csak a reliabilitás számításának egyik alapgondolatát kívánjuk bemutatni.

Képzeljük el, hogy van egy tesztünk, amelyik csupa olyan feladatból áll, amely feladatok mindegyike ugyanazt a tulajdonságot méri. Például készítünk egy tesztet, amelynek minde-

gyik feladata azt vizsgálja, hogy a tanulók hogyan tudnak törtet törttel osztani. Ha a feladatok nagyon hasonlóan egymásra, azaz lényegében csak a számértékek különbözőek, akkor azt várhatjuk, hogy aki az egyik feladatot megoldja, az nagy valószínűséggel megoldja a másik feladatot is, aki pedig egyik feladatot elrontja, az elrontja a másikat is. Természetesen vannak határesetek vagy középső esetek, amikor a tanuló a feladatok egy részét meg tudja oldani, más részét nem. A tipikus azonban mégiscsak az, hogy vagy tud törtet törttel osztani, vagy nem, és ha tudja, akkor a feladatok többségét megoldja, ha nem tudja, akkor a feladatok nagy részét nem oldja meg. Ha tehát a tesztünk jól mér, akkor azt várhatjuk, hogy az egyes feladatok megoldása között magas korrelációt kapunk. Ha azonban a teszt nem megbízható, nem jól mér, például véletlenszerűen, találgatással is helyes választ lehet az egyes kérdésekre adni, akkor az egyes feladatok eredményei között, az egyes feladatok megoldásának jósága között semmiféle összefüggés nem lesz. Ez a gondolatmenet felhasználható a reliabilitás vizsgálatának egyik nagyon egyszerű módjára.

Vegyünk egy tesztet, amely viszonylag sok (10-15) itemből áll. Osszuk két részre a kérdéseket, sorszámozzuk be, és vegyük a páros és páratlan itemeket külön-külön. Képezzünk így egy képzeletbeli féltesztet a páros, és egy másik képzeletbeli féltesztet a páratlan itemekből. Adjuk most össze külön a páros és külön a páratlan számú itemek eredményét, és számítsuk ki a korrelációt az így kapott két félteszt között. Töltsünk ki képzeletben egy tesztet véletlenszerűen, például kockadobással. Ekkor a két félteszt eredménye között semmiféle összefüggés nem lesz: semmi sem befolyásolja azokat olyan irányba, hogy ha az egyik félteszt az átlagosnál jobb, akkor a másik félteszt is az legyen. A két félteszt eredményeit ugyanis csak a tanuló tudása kapcsolhatja össze. Azt várhatjuk, hogy ha a teszt jól méri azt, aminek a mérésére használni akarjuk, akkor azok akik a teszt egyik felét jól oldották meg, jól fogják megoldani a másik felét is, akik az egyik fe-

lét rosszul oldották meg, rosszul oldják meg a másik felét is. Minél szorosabb a két félteszt között a kapcsolat, annál inkább a tanuló tudása határozza meg a teszteredményt, és nem valami más külsődleges tényező, például találgatás.

A reliabilitás vizsgálatára más alapelveket, például a varianciák összehasonlítását is felhasználhatjuk. A gyakorlatban a reliabilitás kiszámításának konkrét módját általában attól tesszük függővé, hogy milyen adatok állnak rendelkezésünkre. A reliabilitás a teszt alapvető jellemzője. A tesztet bárminek a mérésére csak akkor használhatjuk, ha magas reliabilitási mutatókkal rendelkezik.

3.3.3. Validitás

Validitáson a teszt érvényességét értjük: azt a tulajdonságát, hogy a teszttel valóban azt mérjük-e, aminek a mérésére a tesztet kidolgoztuk, amit a teszttel mérni kívánunk. A kérdés megint túlságosan is egyszerűnek tűnik, hiszen ha egy tesztet elkészítettünk valaminek a mérésére, szinte természetesnek tűnik, hogy a teszt valóban azt is méri. Ez azonban egyáltalán nem biztos. A problémát megint egy egyszerű példával szemléltethetjük.

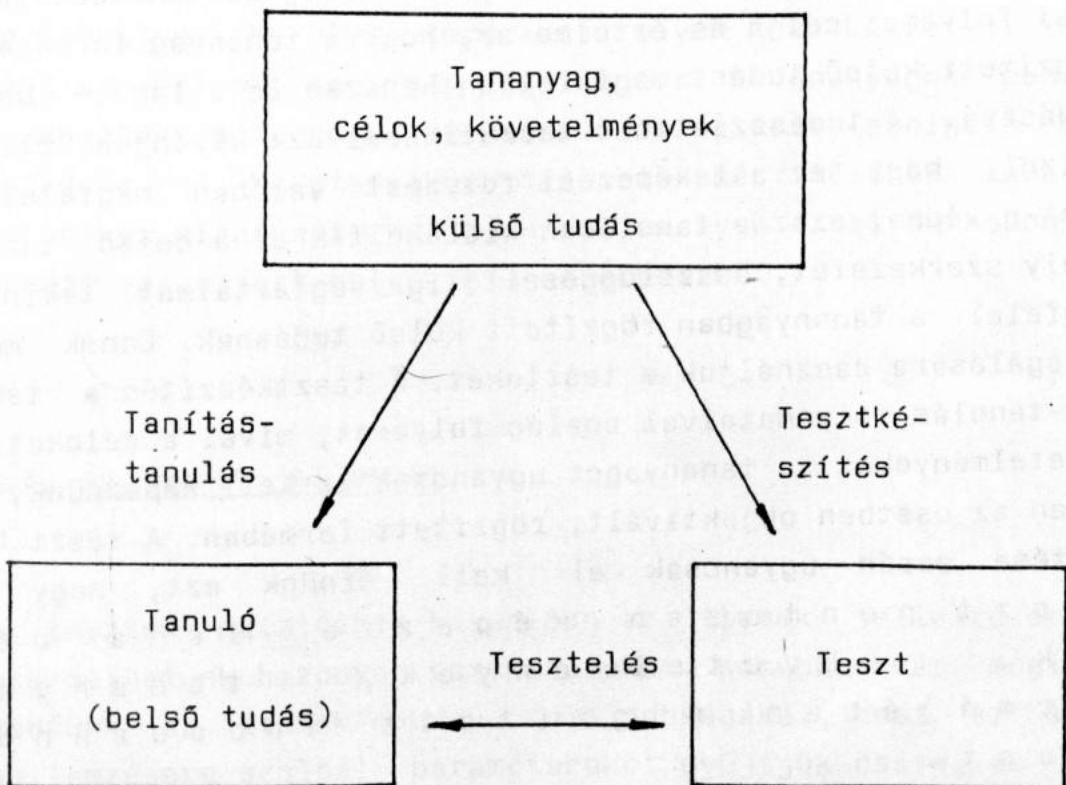
Tételezzük fel, hogy egy tesztet készítünk, amivel az általános iskolai tanulók matematika tudásának egy speciális részét, például a szöveges feladatok megoldásának készségét kívánjuk vizsgálni. Elvégeztük a felméréseket, és az eredmények értékelésénél tartunk. Tegyük fel továbbá, hogy rendelkezésünkre állnak a tanulók korábbi tesztjei, így például az írásbeli szorzás és osztás végzésével kapcsolatos tesztek, és ugyancsak rendelkezésünkre állnak a tanulók olvasásmegértési tesztjeinek eredményei. A kiértékelés során furcsa dolgot ta-

pasztalunk: azt látjuk, hogy azok a gyerekek, akik a matematikában jók, rendszeresen rossz megoldást adnak a teszten, míg mások, akik korábban nem jeleskedtek, egészen jól megoldják a feladatokat. A furcsa jelenség feltárása érdekében számítógépbe tápláljuk a korábbi matematika teszteredményeket, az olvasás teszt eredményeit és a szóban forgó szöveges feladat megoldásával kapcsolatos teszt eredményeit is. Kiszámítjuk a tesztek eredményei közötti korrelációt, és azt tapasztaljuk, hogy a szöveges feladat tesztünk eredményei sokkal szorosabban korrelálnak az olvasás teszt eredményeivel, mint a korábbi matematika eredményekkel. A jelenség oka tehát rögtön egyértelművé válik, kissé pontosabb elemzéssel rá is jövünk, hogy a szöveges feladat teszt nem azt méri, aminek a mérésére kidolgoztuk: a szövege ugyanis túlságosan bonyolult. A tanulók egy része el sem jut a feladat megoldásáig, vagy legalábbis nem annak a feladatnak a megoldásába fog bele, amit mi kitűztünk, mert nem érti meg a feladat szövegét. Olvasási készsége nem elég fejlett ahhoz, hogy a feladat bonyolult szövegét pontosan értelmezze. Hiába tudná a szöveg mögött rejlő, a számára esetleg egyszerű matematikai feladatot megoldani, ha a bonyolult szöveg megakadályozza abban, hogy ezt a feladatot megértse, és a matematikai megoldást megadja.

Tesztünk tehát valóban mér valamit, mégpedig azt, hogy a tanulók mennyire jól értik meg a feladat szövegét, azaz milyen az olvasási készségük. A teszt eredményeiben e készség fejlettsége tükröződik, nem pedig az, aminek a mérésére kidolgoztuk, azaz a szöveges feladatok megoldásának matematikai készségei. Ebben az esetben azt mondjuk, hogy a teszt validitása nem megfelelő, nem azt méri, aminek a mérésére készítettük. Emellett a teszt rendelkezhet még magas reliabilitási mutatókkal, azaz a teszt valamit jól mérhet, csak nem azt, amit mi vele mérni gondoltunk.

Érdemes utalni a feladat reliabilitásának és validitásának összefüggéseire. Ahhoz, hogy a tesztnek a validitásáról egyáltalán beszélhessünk, a tesztnek megbízhatónak kell len-

nie, tehát jó reliabilitási mutatóval kell rendelkeznie. Az összefüggés fordítva nem áll fenn, ha egy tesztnek rossz a validitása, attól még a reliabilitása lehet igen magas. Láttuk, hogy egy teszt validitása még viszonylag egyszerű esetekben is kétséges lehet. Szinte természetes, hogy olyan bonyolult pszichikus tulajdonságoknál, mint például az intelligencia, a kreativitás, a problémamegoldó gondolkodás, a megfelelő tesztek validitása még sokkal nehezebb kérdéseket vet fel.



2. ábra

A tudásszintmérő tesztek validitása

Sok pszichológiai teszt validitása, validitásának igazolása csak igen nehezen megoldható problémát jelent, általában már validnak elfogadott tesztekhez viszonyítják az újonnan kidolgozott tesztek. A pedagógiai tesztek,

közelebből a tudásszintmérő tesztek esetében a kérdés kissé egyszerűbb, vannak olyan támpontjaink és olyan technikai eljárások, amelyek segítségével a teszt készítése során elérhetjük, hogy a teszt megfelelő validitással rendelkezzen.

A tudásszintmérő tesztek validitásának kérdéseit a 2. ábra alapján tekinthetjük át.

Az egyszerűség kedvéért induljunk ki abból az alapmodellből, hogy a célokban, követelményben, tananyagban rögzített külső tudásból a tanítás-tanulás folyamatai során a tanuló belső tudása jön létre, ennek eredményeképpen alakulnak ki a megfelelő pszichikus rendszerek. Az egész tanítási-tanulási folyamat célja és értelme az, hogy a tananyag formájában rögzített külső tudást megfelelően képezze le a tanuló belső tudásává. A tudásszintmérő teszteléssel azt kívánjuk ellenőrizni, hogy ez a leképezési folyamat valóban megfelelően ment-e végbe, azaz a tanulóban kialakult-e az a belső tudás, amely szerkezetét, összefüggéseit, igazságtartalmát tekintve megfelel a tananyagban rögzített külső tudásnak. Ennek megvizsgálására használjuk a tesztekét. A tesztkészítés a tanítás-tanulás folyamataival analóg folyamat, mivel a célokat, a követelményeket, a tananyagot ugyancsak le kell képeznünk, de ebben az esetben objektivált, rögzített formában. A teszt készítése során ugyancsak el kell érünk azt, hogy a teszt pontosan képezze le a célokat, követelményeket, tananyag formájában rögzített tudásrendszert.

Ha rendelkezésünkre áll egy olyan teszt, amelyik pontosan leképezi a tananyagot, akkor a tanulót a tesztelés folyamata során szembesíthetjük ezzel a teszttel és megvizsgálhatjuk, hogy a tanuló pszichikumában keletkezett belső tudás jól képezi-e le a célokat, követelményeket, tananyag formájában rögzített külső tudást. Belátható, hogy ez az összehasonlítás csak akkor végezhető el hitelesen, csak akkor tudjuk a tesztünkkel valóban azt vizsgálni, hogy a tanulás eredményeként

milyen belső tudás jött létre, illetve hogy ez a belső tudás megfelel-e a céloknak, követelményeknek, ha a tesztünk pontosan a célok, követelmények, illetve a tananyag alapján készült, pontosan tükrözi azt. Ha tesztünk ele-
gendően részletes és megbízható, akkor a teszteredmények valóban azt mutatják meg, hogy a tanuló tudása mennyiben egyezik meg a tananyag formájában rögzített tudással.

A tudásszintmérő tesztek validitását tehát azáltal biztosíthatjuk, ha olyan technikát, olyan tesztkészítési eljárást használunk, amelynek segítségével a tananyagot pontosan leképezzük a tananyag mérésére, tesztelésére szolgáló mérő-eszközbe. A következőkben a tesztek készítésének azokat a technikai eljárásait mutatjuk be, amelyek az ilyen módon elérhető validitást helyezik a középpontba.

3.4. Az itemek jellemzői

A tesztek mutatóihoz hasonlóan az itemek jellemzésére is használhatunk bizonyos paramétereket. Korábban már megismertünk az itemek néhány tulajdonságával, most az itemek jellemzésére szolgáló paramétereket gyűjtjük össze, és röviden megadjuk ezek értelmezését.

Az item nehézsége, vagy nehézségi indexe azt mutatja meg, hogy egy tetszőlegesen kiválasztott tanuló milyen valószínűséggel oldja meg az itemet. Gyakorlatilag ez nem más, mint a jó megoldások száma osztva a feladatot megoldó tanulók teljes számával. Az item nehézségi indexe tehát 0 és 1 közé eső szám. Minél nagyobb ez a szám, annál könnyebb az item.

Az item differenciáló ereje vagy megkülönböztető képes-

sége azt mutatja meg, hogy az item mennyire érzékeny a tanulók tudására, mennyire élesen tesz különbséget a különböző tudásszintű tanulók között.

Beszélhetünk az itemek megbízhatóságáról, reliabilitásáról is. Az itemek reliabilitását is sokféle mutatóval jellemezhetjük. A legegyszerűbb mutató, amiből az itemek reliabilitására következtethetünk, az itemnek és a teszt összpontszámának a korrelációja. Könnyű belátni ugyanis, hogy ha egy item jó, akkor azt az itemet azok a tanulók oldják meg jól, akik az egész tesztet is jól megoldják, tehát az item eredményei és a teszt összpontszáma szorosan korrelálnak. Ha ez a korreláció alacsony, az azt jelenti, hogy az item nem azt méri, amit a teszt egésze, vagy nem jól méri azt, amit a teszt egésze. Ha az item - tesztösszpontszám korreláció negatív, akkor az adott itemet tendenciaszerűen azok a tanulók oldották meg jól, akik a tesztet egészében rosszul, és megfordítva, azok a tanulók oldották meg rosszul, akik a teszt egészén többnyire jól szerepeltek. Ilyen esetben arra gyanakodhatunk, hogy az item körül valami hiba van, a megfogalmazás félreérthető. Ezekben az esetekben általában a feladat, vagy az item elemzésével csakhamar megtaláljuk a hibát. Természetesen a már nagyon kis pozitív érték is azt jelenti, hogy az itemmel valami probléma lehet.

3.5. Normaorientált és kritériumorientált tesztelés

A klasszikus tesztelmélet nemcsak matematikai (statisztikai, valószínűségelméleti) modelljeit tekintve fejlődött tovább, hanem szemléletmódjában is, amit az ún. kritériumorientált tesztelés kialakulása jelez. Már az előző részben foglalkoztunk azzal, hogy sok pszichológiai teszt esetében nem tudunk külső kritériumokat találni, amelyek segítségével a mérendő tulajdonság maximális értékét megadhatnánk, és nehézséget o-

koz annak pontos meghatározása is, hogy mit mér a tesztünk. Ezzel szemben a tudásszintmérő teszteknel a cél- és követelményrendszer megfelelő támpontot ad a validitás biztosításához. Ez a különbség vezetett (többek között) a kritériumorientált tesztelés gyors fejlődéséhez.

A kritériumorientált és a klasszikus tesztelés különbségeinek bemutatása érdekében először a klasszikus tesztelmélet és értékelési eljárás egyik lényeges vonását kell tovább elemeznünk. Ahhoz, hogy a teszt összpontszámának, a teszteredménynek valamilyen jelentést tulajdonítsunk, ezt a pontszámot valamihez viszonyítanunk kell. Önmagában például nem tudjuk, mit jelent az, hogy egy tanuló 60 pontot ért el egy teszten. Más ennek a 60 pontnak a jelentése akkor, ha a minta - a tanulóknak az a csoportja, amely a tesztet megoldotta -, 50 pontos átlagteljesítményt ért el. Megint más a helyzet akkor, ha az átlag 70 pont volt. Ebben az esetben a kiszemelt tanulót annak a csoportnak a teljesítményéhez viszonyítottuk, amelyik a tesztet megoldotta. A viszonyítási szint megadásával nyert értelmet a 60 pontos teljesítmény. Ugyanezt az alapelvet megtartva tovább finomíthatjuk eljárásainkat. Nemcsak azt mondhatjuk meg, hogy ez a 60 pont az átlagnál jobb, vagy rosszabb, de egész pontosan megmondhatjuk azt is, hogy a tanulók hány %-a ért el ennél gyengébb vagy jobb eredményt. Így elkészítettünk egy olyan viszonyítási rendszert, amelyben minden egyes megvizsgált személyt a populáció jellemzőihez viszonyítunk. Mivel itt a populáció normáihoz viszonyítunk, ezt a viszonyítási eljárást normaorientált, vagy normavonatközös viszonyításnak nevezzük.

Alkalmazhatunk egy másik viszonyítási eljárást is, ha létezik a vizsgált tulajdonságnak valamilyen természetes maximuma, vagy egyéb természetes viszonyítási pontja. Ha például a tanulóknak egy adott nyelv tanulása során adott időszakban egy bizonyos számú szót kell megtanulniuk, készíthetünk olyan tesztet, amelynek segítségével azt mérjük, hogy a

tanuló ennek a szókincsnek hány %-át birtokolja. Hasonlóképpen készíthetünk tudásszintmérő teszteket, amelyek azt vizsgálják, hogy a tantervben rögzített követelmények hány %-át sajátították el a tanulók. Ebben az esetben a tanulók viszonyításánál semmiféle szerepet nem játszik az, hogy más tanulók milyen eredményt értek el, egyedül csak egy külső, előre rögzített, meghatározott kritériumhoz viszonyítjuk a tanulókat. Ez a fajta viszonyítás egy egészen más szemléleten nyugszik, mint a normaorientált viszonyítási rendszer, ezt a viszonyítási megoldást kritériumorientált tesztelésnek nevezzük.

Bár itt látszólag csupán értékeléstechnikai kérdésekről van szó, a normaorientált és a kritériumorientált értékelés különbözősége nemcsak a viszonyítási szempontban rejlik. Az ugyanis, hogy egy külső kritériumhoz viszonyítjuk az adott tanuló teljesítményét, magában foglal egy további feltételezést is, nevezetesen azt, hogy ilyen külső kritériumokat meg lehet adni, például meg tudjuk adni azt, hogy mit tekintünk a vizsgálandó tulajdonság 100 %-ának. A pszichológiai tesztek és a klasszikus tesztelmélet tehát nem azért használja a norma-vonatkoztatású viszonyítási rendszert, mert ezt tartja a legelőnyösebbnek, hanem azért, mert sok pszichikus tulajdonság tekintetében nem tudunk megadni természetes maximumot. Nem lehet például megadni azt, hogy mennyi az intelligencia 100 %-a, vagy ki a 100%-osan kreatív ember, mi az a maximum, ami fölé nem mehetünk.

A tudásszintmérő tesztekhez lehet maximumokat, külső kritériumokat kitűzni. Ha arra vagyunk kíváncsiak, hogy a tanulók milyen mértékben teljesítették a követelményeket, akkor ezeket a kitűzött követelményeket a tanulók különböző mértékben teljesíthetik. Teljesíthetik 20, 30, 40, 60, vagy akár 100 százalékos mértékben. Ebben az esetben természetesen nem a tanuló teljes, különböző forrásokból megszerzett tudását értékeljük. Lehet, hogy bizonyos elemeiben a tanuló tu-

dása messze meghaladja azokat a követelményeket, amelyeket mi számára kitűztünk. A dolog természeténél fogva ilyenkor is csak azt tudjuk a teszt eredményei alapján mondani, hogy a tantervi követelményeknek 100 százalékosan megfelel.

A kritériumorientált tesztelés mindezeknek megfelelően sajátos tesztkészítési technikákat igényel. E teszt-készítési technikák középpontja, alapja a tananyag megfelelő elemzése. Csak a tananyag elemzése, szerkezetének pontos feltárása révén készíthetünk megfelelő teszteket, vagyis csak így érhetjük el azt, hogy tesztünk pontosan leképezze a tananyagot. A megfelelő teszt készítéséhez azonban nemcsak a tananyagot kell ismernünk, hanem pontos képpel kell rendelkezünk arról is, hogy milyen természetű az a tudás, milyen jellegűek azok a rendszerek, amelyek a tanulók pszichikumában kialakulnak. A tesztelés során ugyanis nem csupán a tanuló által elsajátított külső tudás meglétét vizsgáljuk, hanem egy egészen más minőséget, pszichikus tulajdonságaikat, pszichikus szabályozási rendszereket, amelyek alapvetően más sajátosságokkal rendelkeznek, mint az egyébként elemezhető, tankönyvek, tananyagok formájában megjelenő külső tudás.

4. F E L A D A T Í R Á S

4.1. A tananyag elemzése, a felmérendő tudás kategorizálása

Egy-egy tesztfeladattal, teszt-ítemmel az elsajátított tudásnak csak meglehetősen kis elemeit tudjuk vizsgálni. Ahhoz, hogy a tudásnak már egy ilyen kis eleméhez is hozzáférjünk, és ugyanakkor a tudás elemenkénti megközelítésével a vizsgálandó tudás teljes épületét is megfelelően építsük fel, szükség van arra, hogy a tananyagot megfelelő módon elemezzük.

Mind a tananyag elemzésére, mind a tananyag elsajátítása nyomán kialakuló pszichikus rendszerek számbavételére sokféle megoldás született, sokféle technológia ismeretes. Az oktatás céljainak, és így az értékelés objektumainak a számbavételére is alkalmazható egyik legelterjedtebb taxonómiai rendszer Benjamin Bloom nevéhez fűződik. Bloom, majd később munkatársai értékelési taxonómiájukat a személyiség három nagy területére, a kognitív, az affektív és a pszichomotoros szférára dolgozták ki.

A bennünket közelebbről érdeklő kognitív taxonómia rendszerezte az elsajátított tudás különböző elemeit, és ezek között az elemek között hierarchikus szinteket állapított meg. Bloom az összegyűjtött elemeket 6 hierarchikus szintbe foglalta össze. Bár Bloom elmélete a maga idejében igen nagy hatású volt, mára már sokan bírálják a koncepcionális megalapozás hiányát, a taxonómiába felvett elemek esetlegességét. A Bloom által feltételezett hierarchiát sok vizsgálat nem igazolta. Más hasonló, vagy bonyolultabb modellek is születtek, amelyek ismertetésére nem vállalkozhatunk. A következőkben a pedagógia elterjedt szóhasználatánál maradván egy leegyszerűsített megoldást vázolunk fel.

A tanulók által elsajátított, a tesztekkel megvizsgálán-

dó tudást két fő szférára osztjuk fel. Az egyikbe tartoznak az ismeret-jellegű tudáselemek, így például a képzetek, a fogalmak, a tények, a definíciók, a szabályok, a leírások, a törvények és az elméletek. Ezekre a tudáselemekre, különösen a képzetekre, tényekre, egyszerű fogalmakra, definíciókra az jellemző, hogy azok viszonylag rövid idő alatt elsajátíthatók, megfelelő számú ismétlés és rögzítés után tartós tudássá válhatnak. A tudásnak ebbe a szférájába tartoznak például a következő tények: a vas vegyjele Fe, a mohácsi vész 1526-ban volt, az asztal németül der Tisch stb.

A tudás másik nagy csoportjába, a képesség-jellegű tudás körébe tartoznak a készségek, a jártasságok és az általános képességek. Készség például az íráskészség, az olvasás elemi készségei, az alapműveleti számolási készségek. Jártasságot alakíthatunk ki például a tananyag egy meghatározott témakörében való feladatmegoldásra, így a gáztörvényekkel kapcsolatos számításokra vagy a nyelvtanban a mondatelemzésre, vagy jártassá válhatunk szövegek idegen nyelvre történő fordításában. Általános képesség például a problémamegoldó gondolkodás, az intelligencia, a kreativitás, vagy a kommunikáció különböző képességrendszerai.

A képesség-jellegű tudást az jellemzi, hogy hosszú fejlődési folyamaton keresztül jut el a kezdetektől egy viszonylag állandósult szintre. Gondoljunk például az íráskészségre, amelynek elemeit, az írásmozgás-koordinációt már az óvodában elkezdik a gyermekek tanulni, a betűírást az első osztály végére megtanulják, tűrhető sebességű írással rendelkeznek az általános iskola végére, de az állandósulás jegyei csak a kamaszkor vége felé jelentkeznek, és a kiírt írás, amely állandóvá válik, csak a felnőttkor küszöbén jelenik meg. Számolási készségeink is viszonylag hosszú időn keresztül fejlődnek és alakulnak ki, bizonyos ál-

talános képességeink pedig esetleg egész életünkön keresztül is fejlődhetnek. Ezeknek a képesség-jellegű tudáselemeknek az elsajátítása többnyire nem rendelhető az oktatási folyamat egy viszonylag rövid szakaszához, mint ahogy azt bizonyos ismeretek esetében megtehetjük.

Az ismeret-jellegű és a képesség-jellegű tudás még sok más sajátosságban is különbözik, amely a tesztelés szempontjából fontos lehet számunkra. Ezek közül itt most csak néhányat sorolunk fel.

Az ismeret-jellegű tudás (mivel annak elsajátítása az oktatási folyamat egy rövidebb szakaszán történik) felmérése többnyire egy r ö v i d e b b i d ő s z a k a s z eredményeiről ad képet. Ezzel szemben bármikor is történik a képesség-jellegű tudás vizsgálata, annak eredményében mindig az oktatási folyamat egy h o s s z a b b s z a k a s z a van jelen. Előfordul például, hogy a 8. osztályosok olvasási színvonalát mérjük fel. Ekkor természetesen nem a megelőző 1-2 hónap iskolai munkájának eredményességét vizsgáljuk, hanem a megelőző 6-8 év eredményei tükröződnek az olvasás-szint felmérésben. Így tehát mások lehetnek a teszt eredményeiből levont következtetések, az azok alapján megfogalmazható feladatok is. A tanulók képességeiben talált hiányosságokat csak hosszú, néha több éves kemény munkával lehet pótolni.

Fontos különbség az ismeret-jellegű és a képesség-jellegű tudás között az is, hogy a z i s m e r e t - j e l l e g ű t u d á s t , a n n a k e l e m e i t a m a g u k k o n k r é t f o r m á j á b a n t e h e t j ü k a t e s z t e l é s t á r g y á v á . Például konkrétan rákérdezhetünk, hogy mi a vas vegyjele, mikor volt a mohácsi vész stb. Ezzel szemben a képesség-jellegű tudást a működés felől, a szerkezet felől közelíthetjük meg. Például az olvasás készségét többféle tartalommal is vizsgálhatjuk, sokféle szöveget olvastathatunk az olvasási készség felméréseként. Hasonlóképpen különböző szövegekkel vizsgálhatjuk az írás-

készséget, fizika feladatokat különböző számértékekkel adhatunk és így tovább. A képesség - jellegű tudásnak tehát a szerkezetbeli és működésbeli sajátosságait kell leírniunk és elemeznünk.

A tudás e két formája többnyire meghatározza az alkalmazható teszt szerkesztési technikát, a tesztelésre használható feladattípusokat, tesztelési eljárásokat.

4.2. Feladattípusok

A teszt feladatok készítésének óriási irodalma van. A mérés céljaitól függően sokféle tesztet és teszt feladatot lehet készíteni. Ezek közül a következőkben csak olyan feladatokkal foglalkozunk, amelyek a tudásszint mérésére alkalmasak, még szűkebben csak azokkal, amelyekből papír-ceruza tesztek lehet összeállítani. Természetesen még ezen a csoporton belül is csak vázlatos áttekintésre vállalkozhatunk.

A tudásszint mérésére szolgáló feladatok kiindulásként két nagy csoportra oszthatók, az egyik csoportba a feleletválasztó, a másikba a feleletalkotó feladatokat soroljuk.

Mindkét feladattípusnak megvannak a maga relatív előnyei és hátrányai, továbbá nem egyformán alkalmazhatók a tudás különböző komponenseinek a mérésére.

Az iskolai használat során a feleletválasztó feladatok néhány típusának nagy hátránya az, hogy hamis, nem igaz, ámde hihető válaszalternatívák nagy tömegének az elolvasására készítik a tanulót, és gyenge, bizonytalan tudás esetén van bizonyos valószínűsége a hibás válaszok rögzítésének. Ezért pedagógiai szempontból a hibás válaszokat nem tartalmazó típusokat illetve az aktív válaszadást előnyben részesítjük.

4.2.1. Feleletválasztó feladatok

A feleletválasztó kategóriába azokat a feladatokat soroljuk, amelyeknek a megoldása során a tanuló kész, előre megadott válaszok közül választja ki a megfelelőt (helyeset, jót, igazat stb.), és azt valamilyen módon megjelöli. A feleletválasztó feladatok javítása a legegyszerűbb, kidolgozásuk azonban általában hosszabb időt vesz igénybe. A nehézségek különösen a többszörös választást tartalmazó feladatoknál jelentkeznek, mivel formailag azonos, a megfelelő kontextusba egyaránt illő (egyaránt hihető, a szükséges tudással nem rendelkezők számára jónak tűnő) válaszalternatívákat kell kidolgozni. A feladatok értékelő rendszerének, javítókulcsának és pontozásának elkészítése viszont könnyű.

A feleletválasztó feladatok elsősorban a tudás egyszerűbb elemeinek, az ismeret-jellegű tudásnak a mérésére alkalmasak. Kevésbé, vagy csak igen elmélyült tesztkészítés révén használhatók a képességeknek, a gondolkodás rugalmasságának, a tudás minőségének, a megértésnek, a szintézisnek a vizsgálatára. Ez azonban nem zárja ki azt, hogy akár az általános képességek vizsgálatára is feleletválasztó tesztet szerkesztünk.

A feleletválasztó tesztek sokféle típusa közül itt csak néhányat mutatunk be. (A közölt minták csak az illusztrációt szolgálják, egy részük kitalált példa, más részük a JATE Pedagógiai Tanszékén készült tesztekől származik.) A helytel való takarékoskodás érdekében csak a feladatok lényeges elemeit emeljük ki, tehát nem tüntetjük fel mindazokat a formai tartozékokat, amelyekkel egy valódi tesztbe illesztett feladatnak rendelkeznie kell (ezeket ld. az 5. fejezetben).

4.2.1.1. Alternatív választás

Ennél a feladattípusnál csak két válaszlehetőség adott, és ezek közül az egyik mindig jó, a másik rossz megoldás. A véletlen találat valószínűsége itt a legmagasabb, 50 %. Fontosabb típusai a következők:

(1) Állítás, amelyről el kell dönteni, hogy igaz, vagy hamis. Ha egy zárt térben levő gáz hőmérsékletét növeljük, a nyomása csökken.

igaz hamis

(2) Feladat, amelyről el kell dönteni, hogy a megadott megoldás jó-e. Például jó-e az alábbi angol mondat?

The boy have a book.

jó rossz

(3) Eldöntendő kérdést tartalmazó kérdőmondat.

Voltak-e Mátyás király híres könyvtárában Amerikáról szóló útleírások?

igen nem

4.2.1.2. Többszörös választás

A feladatban több válaszlehetőség van megadva, ezek közül lehet egy, de több is helyes. A feladatnak mindig közölnie kell, hogy egy vagy több választ kell a teszt megoldójának megjelölnie. A véletlen találat valószínűségét matematikailag a válaszlehetőségek száma és a megjelölendő válaszok száma határozza meg, gyakorlatilag az is számít, hogy a válaszlehetőségek mennyire "hasonlítanak" egymásra. Ha ugyanis egy rossz válaszról a vizsgálandó tudás birtoklása nélkül is nyilvánvaló, hogy rossz, akkor valójában csak a többi jön számításba, mint reális válaszalternatíva. Ezzel a

megoldással nagyon sokféle feladatot lehet szerkeszteni, itt csak néhány alaptípust sorolunk fel.

(1) Egy jó válasz

Mi az Amerikai Egyesült Államok fővárosa?

- (a) Montreal
- (b) New York
- (c) Washington
- (d) Los Angeles

c

(2) Több jó válasz

Kik voltak az első triumvirátus tagjai?

- (a) Antóniusz
- (b) Cézár
- (c) Lepidusz
- (d) Krasszus
- (e) Pompejus
- (f) Oktáviusz

b, d, e

(3) A legjobb válasz kiválasztása

Melyik szó adja vissza legjobban magyarul a "kompetencia" jelentését?

- (a) tudás
- (b) szaktudás
- (c) műveltség
- (d) szakértelem
- (e) képzettség

d

(4) A rossz válasz kiválasztása

Melyik az a város, amelyik nem az Egyesült Államokban van?

- (a) Boston
- (b) Chicago
- (c) San Francisco
- (d) Canberra
- (e) Santa Barbara

d

4.2.1.3. Válaszok illesztése

Ennél a kérdéstípusnál két halmaz elemei között kell a kapcsolatot megtalálni. A két halmaz nagyon sokféle módon állhat tartalmi kapcsolatban egymással, például dolgok és tulajdonságaik, szerzők és művek, fogalmak és általánosabb fogalmak stb. A hozzárendelés típusa lehet egy az egyhez és egy a többhöz (ritkábban több a többhöz is). Természetesen a feladatban mindig pontosan rögzíteni kell, hogy hogyan várjuk a választ.

(1) Egy az egyhez hozzárendelés.

Kapcsoljuk össze a szerzőket és a műveket!

- | | |
|-----------------------------|---------------------|
| (a) A fekete város | (A) Ady Endre |
| (b) A fekete kolostor | (B) Mikszáth Kálmán |
| (c) A fekete zongora | (C) Kuncz Aladár |
| a - B b - C c - A | |

(2) Egy a többhöz hozzárendelés (osztályozás).

Melyik szó melyik szófajhoz tartozik?

- | | |
|---------------|-------------|
| (a) ige | (A) Petőfi |
| | (B) alszik |
| (b) főnév | (C) kicsi |
| | (D) sok |
| (c) melléknév | (E) magas |
| | (F) enyém |
| (d) számnév | (G) olvas |
| | (H) írás |
| (e) névmás | (I) szépség |

a: B, G b: A, H, I c: C, E d: D e: F

A feleletválasztó feladatoknak a felsoroltakon kívül még nagyon sok változata használatos. Mivel a számítógépes tesz-

telést is egyszerűbb feletválasztó feladatokkal megoldani, a számítógépes oktatás korai szakaszában ez a forma tejedt el. Ma azonban, a programok "intelligensebbé" válásával már a számítógépek is képesek a tanulók által alkotott bonyolult válaszokat is kiértékelni.

4.2.2. Feleletalkotó feladatok

A feleletalkotó feladatok megszerkesztése viszonylag könnyű munkának tűnik, hiszen látszólag általában elegendő egy felszólító vagy kérdőmondatot megfogalmazni. Azonban e feladatok értékelő rendszerének megalkotása, az egyértelmű, kellő objektivitással használható javítókulcs kidolgozása többnyire időigényes feladat.

A feleletalkotó feladatokkal a tudás bonyolultabb összetevőinek a mérését is megoldhatjuk, azonban ez gyakran speciális értékelő eljárást is igényel. Kellő gyakorlattal azonban a feleletalkotó feladatok javítókulcsát annyira egyértelművé lehet tenni, hogy az elfogadható válaszlehetőségek, illetve az azokból értékelhető mozzanatok olyan kategóriákba sorolhatók be, amilyeneket a feleletválasztó feladatok esetében is kapunk.

4.2.2.1. Kiegészítés

Ez a feladattípus többnyire egy kijelentő mondat, amelyből bizonyos fogalmat vagy fogalmakat kihagytunk.

(1) Egy kihagyott fogalom.

Az cselekvést, történést, létezést jelentő szó.

(ige)

(2) Több kihagyott fogalom.

A forgatónyomaték megegyezik az és az szorzatával.

(erő, erőkar)

4.2.2.2. Rövid válasz

Ennél a feladattípusnál egy kiegészítendő kérdést úgy fogalmazunk meg, hogy arra röviden, lehetőleg egyetlen szóval, névvel vagy számmal lehessen válaszolni. Értékelése teljesen egyértelművé tehető, ha a kérdést úgy fogalmazzuk meg, hogy csak egyetlen jó megoldást lehessen adni.

(1) Válasz tulajdonnévvel

Ki a Csongor és Tünde szerzője?

(Vörösmarty Mihály)

(2) Válasz számmal

Mikor volt a Mohácsi csata?

(1526-ban)

(3) Válasz egyéb szóval

Melyik névelő tartozik a Tisch főnévhez?

(der)

4.2.2.3. Hosszú válasz

A feladat olyan kérdést tartalmaz, amely hosszabb, általában egész mondatos választ vagy egy felsorolást igényel. A válasz értékelése bonyolultabb, mivel ugyanazt a tartalmat többféle mondattal is meg lehet fogalmazni. Ekkor azt kell megítélnünk, hogy a válasz kifejezi-e a lényegét. Ezt pedig úgy tehetjük meg, hogy megadjuk azokat a kifejezéseket, amelyeknek elő kell fordulniuk a válaszban, és azt, hogy ezeknek

milyen kapcsolatban kell állniuk egymással.

(1) Válasz egy mondattal (definíció).

Mi az ige?

(szó; és cselekvést, vagy történést, vagy létezést jelent)

(2) Válasz felsorolással.

Melyek a halogén elemek?

(fluór, klór, bróm, jód, asztácium)

4.2.2.4. Esszé típusú válasz

A legbonyolultabb értékelési problémákat az egy mondatnál hosszabb összefüggő válaszokat igénylő feladatok jelentik. Az ilyen feladatok alkalmasak az önálló ítéletalkotás, az összefüggések felismerésének, a szintetizálás, a lényegkiemelés képességének a vizsgálatára. Az értékelés objektivitását általában azoknak az állításoknak a felsorolásával lehet elérni, amelyeknek elő kell fordulni a válaszban. Ilyen kérdések lehetnek a következők:

(1) Elemezze a mohácsi csatavesztés okait!

(2) Foglalja össze a kénsavgyártás folyamatait!

4.2.3. A képesség-jellegű tudás mérésére alkalmas feladatok

A korábban bemutatott feladattípusok elsősorban az ismeret-jellegű tudás vizsgálatára alkalmasak. Természetesen megfelelő technikával néhány típusuk a képességek mérésére is alkalmas, az ismeret- és a képesség-jellegű tudás közötti különbségek miatt azonban szükség van más, speciális feladattípusok

alkalmazására is. A képesség-jellegű tudás vizsgálatára szolgáló feladatok nagyon sokfélék lehetnek, néha egészen egyszerűek, máskor bonyolult értékelési technikákra van szükség. Ezeknek a feladattípusoknak a részletes bemutatására terjedelmi okok miatt nem vállalkozhatunk, de az érdeklődő olvasó gazdag példatárat találhat az ajánlott irodalomként felsorolt munkákban.

Az egyszerű alpműveleti számolási készségek tesztelésére jól használhatók azok a feladatok, amelyeket az iskolai gyakorlatban egyébként is használunk. Többnyire elegendő kitűzni magát a feladatot, az eredményről egyértelműen eldönthető, hogy jó vagy rossz. Például:

$$27+16=... \quad 48-17=... \quad 33+...=51$$

Hasonlóképpen könnyen teszté alakíthatjuk a szöveges feladatok megoldásának jártasságait felmérő szokásos iskolai feladatokat. Mivel az ilyen feladatok már több önállóan is értékelhető részfeladatból állnak, mindig pontosan meg kell határoznunk a feladatot felépítő itemeket és az azokra adható pontszámokat. Ilyen jellegű feladatot lehet készíteni a fizikában és a kémiában kialakítandó feladatmegoldó jártasságok vizsgálatára is.

Speciális technikákat igényel az anyanyelvi kommunikáció képességrendszerének vizsgálata. Az írás készségeit az egyszerű írásmozgás-koordinációtól kezdve a helyesírás és mondatalkotás készségein és jártasságain keresztül a fogalmazástechnika képességeiig vizsgálhatjuk, és erre ma már sokféle tesztelési technika áll rendelkezésünkre.

A képességek kialakulása hosszú fejlődési folyamat eredménye. Amikor a képességeket (készségeket, jártasságokat) vizsgáljuk, azt akarjuk megállapítani, hogy az egyes tanulók ebben a fejlődési folyamatban hol tartanak. Ezért a képesség-jellegű tudás fejlettségét két olyan mutatóval is jellemezhetjük, amelynek az ismeret-jellegű tudással kapcsolatban nincs lényeges szerepe: beszélhetünk a képességek (készségek, jártasságok) működésének a sebességéről és hibátlanóságáról.

A készségek és jártasságok sebessége a kezdeti megjelenésüktől az automatizáció és a begyakorlottság növekedésével a teljes kifejlődésig egyre fokozódik. A sebességet általában az időegység alatt megoldott feladatok száma jellemezzük. Például az alpműveleti számolási készségek esetében használhatjuk a művelet/perc dimenziót.

A képességek működésének fejlődését, összehangolódását az is mutatja, hogy csökken a hibázások száma. A hibátlanossági arányt jellemezhetjük például azzal, hogy a jól megoldott feladatok számát az összes megoldott feladatok számához viszonyítjuk.

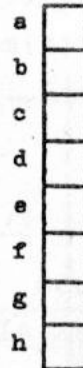
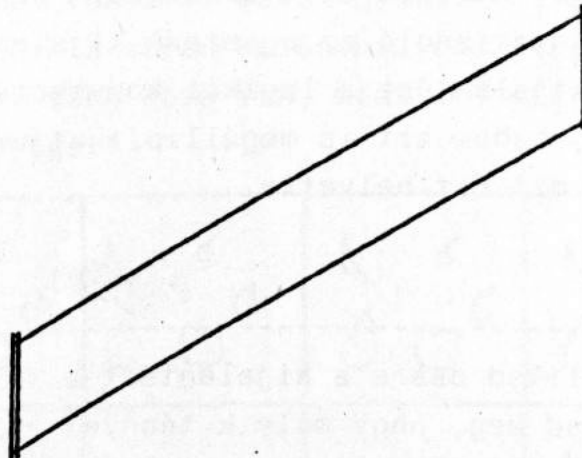
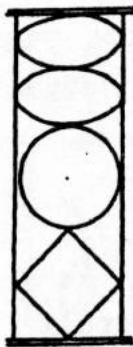
A készségek és a képességek esetében mind a sebesség, mind pedig a hibátlanossági arány (minőség) tekintetében használhatjuk külső viszonyítási pontként (kritériumként) a felnőtt szakember megfelelő mutatóit.

A képesség-jellegű tudás mérése néha egészen speciális megoldásokat igényel. Ezekre a speciális technikákra mutatunk be három példát.

Az alábbi feladat a rajzkészség egyik összetevőjének, a térszemléletnek az elemeit vizsgálja. A feladat megoldásához szükséges transzformációs műveletek, a szem-kéz koordináció, a kézügyesség természetesen nemcsak a rajzolásban játszanak szerepet.

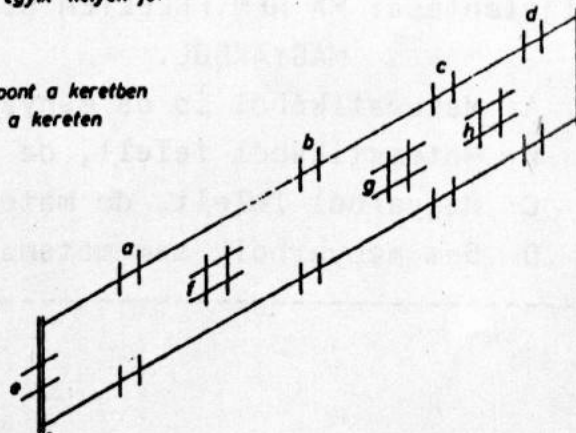
A feladat értékelése a 3. ábrán látható átlátszó fóliával történik, melyet a tanuló rajzára helyezünk, és leolvassuk, hogy megoldása mennyire pontos. Ha az egyes érintkezési pontok a megadott intervallumba esnek, mindegyikre egy pontot lehet kapni.

3. A bal oldali rajz két pálcika közé kifeszített gumihártyát ábrázol. Rajzold be a jobb oldali ábrába, milyen lesz a gumihártyán levő rajz, ha azt az ábra szerinti alakúra nyújtjuk!



3.

a.	}	2 p., ha alul is és felül is jó
b.		1 p., ha csak egyik helyen jó
c.		
d.	}	1 p., ha jó.
e.		
f.	}	2 p., találkozáspont a keretben
g.		1 p., áthaladás a kereten
h.		



3. ábra

Egy térszemlélet-teszt feladat javítására használható fólia rajza

A következő feladat a gondolkodás egyik műveleti képességének, a nyelvi-logikai műveletek kialakultságának a vizsgálatára szolgál. Nemcsak arra alkalmas, hogy segítségével eldöntsük, vajon a tanuló az egyszerű kijelentésekből alkotott összetett kijelentést a logikai konvencióknak megfelelően értelmezi-e, hanem azt is megállapíthatjuk, hogy ha nem így érti, akkor mit ért helyette.

Gondosan hasonlítsd össze a kijelentést a felsorolt tényekkel, és állapítsd meg, hogy melyik ténnyel való összehasonlításban igaz, melyikben hamis a kijelentés!

Minden tényt egyenként hasonlíts össze a kijelentéssel. Amennyiben igaznak találod, a tény előtti betűjelzést karikázd be! Ha megítélésed szerint hamis, akkor a tény előtti betűjelzést húzd át!

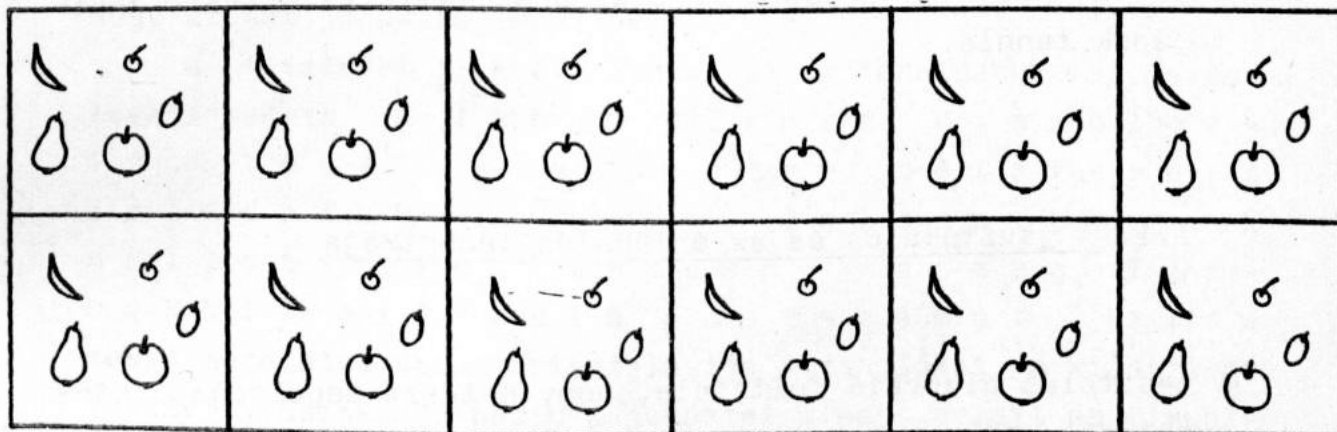
János kijelentése: MA NEM FELELTEM SEM MATEMATIKÁBÓL, SEM
MAGYARBÓL.

- Tények:
- A Matematikából is és magyarból is felelt.
 - B Matematikából felelt, de magyarból nem.
 - C Magyarból felelt, de matematikából nem.
 - D Sem magyarból, sem matematikából nem felelt.
-

Az alábbi feladat egy másik műveleti képesség, a kombinatorív képesség értékelésére alkalmas. A tanuló teljesítményének jellemzésére alkalmazhatunk mennyiségi mutatókat (hány konstrukciót alkotott meg, ezek közül hány a jó, hány a feleslegesen ismétlődő), és jellemezhetjük a gondolkodás minőségét is annak elemzésével, hogy a felsorolás során a tanuló

milyen gondolkodási stratégiát használt.

Karikázz be egy zárt vonallal három gyümölcsöt minden lehetséges módon! Minden kis ábrán különböző összeállításokat keretezz be! Vigyázz, több ábra van, mint amennyi különböző lehetőséget találni fogsz!



5. T E S Z T S Z E R K E S Z T É S

5.1. Az értékelési rendszer kialakítása

Ahhoz, hogy egy teszt betölthesse funkcióit, segítségével a

tanulók tudását objektíven és megbízhatóan felmérhessük, szükség van megfelelő értékelő rendszerre is. Az értékelő rendszer kidolgozásának vannak formai-technikai követelményei, amelyeket azonban nem kezelhetünk részletkérdésként, gyakran éppen ezek képezik a megbízhatóság és az objektivitás alapját. A gyakorlatlan tesztszerkesztők által elkövetett leggyakoribb hibák közé tartozik az, hogy nem írják elő egyértelműen, mit kell a teszt megoldójának, illetve a teszt javítójának tennie.

5.1.1. A javítókulcs és az értékelés technikája

A tesztelés alapvető feltétele, hogy a teszt megoldója pontosan tudja, hogy mit kell tennie, milyen módon kell a feladatot megoldania, a kérdést megválaszolni. E z é r t m i n - d e n t e s z t n e k , v a g y h a a t e s z t b e n k ü l ö n b ö z ő t í p u s ú f e l a d a t o k v a n - n a k , a k k o r m i n d e n k é r d é s n e k v a g y f e l a d a t n a k p o n t o s a n t a r t a l m a z n i a k e l l , h o g y m i a z , a m i t a m e g o l - d ó t ó l e l v á r u n k . Ügyelni kell a kérdések megfogalmazására, helyes ha a kérdés nyelvtani formája is kérdőmondat, és pontosan kifejezi, hogy eldöntendő, vagy kiegészítendő kérdésről van szó. Ha válaszalternatívákat adunk, pontosan közölni kell, hogy mi a teendő, egy vagy több jó válasz van, a jót kell bekarikázni, vagy a rosszat áthúzni. Leghelyesebb, ha mindegyiket, vagyis a jók bekerikázását és a rosszak áthúzását is megköveteljük, így mód van annak megkülönböztetésére is, hogy a tanuló rosszul válaszolt, vagy hozzá sem fogott a feladat megoldásához.

Hasonló szigorúsággal kell előírni a tesztet javító pedagógusok feladatát is. Erre a célra a t e s z t e k -

hez külön javítókulcsokat kell készítenünk, amelyek leírják a javítás módját, és megadják a jó megoldásokat. A feladat típusától függően közlik a jó megoldás kódját, vagy megadják azokat a szavakat, tényeket, amelyeknek a válaszban elő kell fordulniuk ahhoz, hogy azokat jónak fogadhassuk el.

A teszteket el kell látnunk az értékeléshez szükséges tartozékokkal. A tudásmérésére szolgáló tesztek esetében célszerű a feladatok vagy feladatelemek mellett feltüntetni a megoldásukkal elérhető pontszámot. Ez már a teszt megoldása során orientálhatja a tanulókat, a kijavított tesztet visszakapva pedig tájékoztat a megszerzett és elvesztett pontokról. Mindemellett a javítást is egyszerűvé teszi. Jól bevált az a megoldás, hogy a feladat mellett a lap szélén külön kódkockák tartalmazzák az elérhető pontokat, és javításkor áthúzzuk azt, amit a tanuló nem ért el. Így a teszt összpontszámának megállapítását az át nem húzott pontok összeadásával végezhetjük el.

5.1.2. Pontozás, súlyozás

A tanulók teljesítményét, a feladatok megoldása során végzett munkáját a lehető legkisebb, még önállóan értékelhető egységekre kell bontanunk. Ezzel nemcsak a javítás technikáját tehetjük egyértelművé, de fontos visszacsatoló funkciója is van: visszakapva a kijavított tesztet, a tanuló pontosan tudja, mi az, amit jól oldott meg, és mi az, amit nem.

A legkisebb önállóan értékelhető feladatelemek (alternatív egységek, itemek) már csak jók vagy rosszak lehetnek. Nem lehet tehát azokra részpontokat adni, ha ugyanis ezeken belül még fokozatokat tudunk megállapítani, akkor ez azt jelenti,

hogy az adott feladatelemet még tovább lehet bontani, és akkor ezt meg is kell tennünk.

További kérdés az, hogy az egyes feladatelemekre hány pontot adjunk. Azt az eljárást, amelynek során a feladatelemekhez különböző pontszámokat rendelünk, súlyozásnak nevezzük, a feladatelemekhez tartozó pontszám a feladatelem súlya.

Nyilvánvaló követelmény, hogy ha a tudásnak vannak fontosabb és kevésbé fontos elemei, akkor egy fontosabb elemnek a tudásával vagy nem tudásával több pontot lehessen szerezni illetve veszíteni, mint egy jelentéktelenebb elemmel. Azt is elvárhatjuk, hogy egy nehezebb feladat megoldásával több pontot lehessen szerezni, mint egy könnyebbel. De hogyan lehet megállapítani, hogy melyik elem fontosabb, melyik nehezebb? A kérdés megválaszolására többféle megoldás is létezik, elméletileg azonban egyik sem teljesen kifogástalan. Mindegyik tartalmaz valamilyen szubjektív vagy önkényes mozzanatot, ez azonban gyakorlati szempontból többnyire nem okoz problémát.

A tudáselemek relatív fontosságának meghatározására a szakértők közmegegyezése szolgálhat: valamilyen módon meg kell szavaztatni a szakembereket (esetünkben a szaktanárokat), hogy melyik tudáselemet mennyire tartják fontosnak. Ez történhet például úgy, hogy megkérjük, rendezzék sorba az elemeket fontosságuk szerint (keves elemnél), vagy értékeljék az elemek fontosságát egy (például öt vagy tíz fokozatú) skálával. Kellő számú szaktanár véleményét begyűjtve, statisztikai átlagolással állapíthatjuk meg az elemek egymáshoz viszonyított súlyát.

A feladatelemek nehézségének megállapítását empirikusan is elvégezhetjük: kellő számú tanulóval megoldatjuk a tesztet, majd (ha a feladatok jóságát egyébként már bizonyítottuk), a feladatot annál nehezebbnek tekintjük, minél kevesebb tanuló oldotta meg.

Még mindig kérdés marad, hogy mekkora a legfontosabb és a legkevésbé fontos elemek pontszámainak az aránya. Itt is csak gyakorlati megfontolásokból indulhatunk ki (amit való-

színűségi számításokkal is alátámaszthatunk): nem lenne helyes, ha egyetlen feladattal túlságosan sokat lehetne nyerni vagy veszíteni. Praktikusan tehát azt lehet ajánlani, hogy ha az itemek legkisebb pontértéke az 1 pont, akkor a legmagasabb súlya se legyen 5-nél nagyobb. Célszerű, ha ez a terjedelem már magában foglalja a fontosságból és a nehézségből fakadó különbségeket is.

Ha egy pedagógus saját használatra készít tesztek, az előző szempontokat mérlegelve alakíthatja ki a feladatok súlyát. A feladatelemek fontosságának megítélésében csak saját szakmai kompetenciájára támaszkodhat, a nehézség mérlegeléséhez felhasználhatja a tanulók korábbi eredményeit.

Ugyancsak a súlyozással függ össze az is, hogy néha kényelmi szempontból célszerű a pontszámokat úgy megállapítani, hogy a teljes tesztre összesen 100 pontot lehessen kapni. Ez azt jelenti, hogy az eredeti (nyers-) pontokat át kell számítani százalékponttra. A számítást egyszerűen elvégezhetjük, ha minden egyes nyerspontot elosztunk a teszt összpontszámával, és megszorozunk százszal. Így viszont többnyire nem egész számot, hanem tizedes törtet kapunk az egyes itemek százalékpontjaként. Ezen ismét úgy segíthetünk, hogy a pontszámokat kerekítjük, de olyan módon, hogy összességében a teszt összpontszáma 100 maradjon. Látható, hogy a százalékponttra való áttérés kismértékben megváltoztatja a feladatok egymáshoz viszonyított súlyát. Ezért mindig a gyakorlati szempontok alapján kell eldöntenünk, mit tartunk fontosabbnak, a kifejezőbb (az összehasonlítás lehetőségét is kínáló) százalékpontot vagy a pontosabb súlyozást.

5.2. Ekvivalens tesztváltozatok készítése

Több elméleti és gyakorlati szempont is indokolja, hogy ugyanannak az átfogó tudásterületnek a felmérésére azonos módon használható, egymással egyenértékű tesztek álljanak rendelkezésünkre. Például célszerű, ha az egymás mellett ülő tanulók nem pontosan ugyanazokat a feladatokat oldják meg, vagy esetleg szükség van arra, hogy ugyanannak a tanulócsoporthoz a tudását egy adott témakörben többször is megvizsgáljuk. Ilyen esetekben használhatjuk az ekvivalens tesztváltozatokat.

Más szempont is indokolja, hogy egy adott témakörben több ekvivalens tesztváltozatot is összeállítsunk. A tudásszintmérő tesztek készítése során arra törekszünk, hogy egy témakör teljes anyagát tesztfeladattokká fogalmazzunk át. Ez rendszerint azt jelenti, hogy több feladatunk van, mint amennyit egy tanulónak feladhatunk, hiszen nincs annyi idő a tesztelésre, hogy mindenkitől mindent kikérdezzünk. De egyben azt is szeretnénk, hogy a teljes tananyag elsajátításának színvonaláról képet kapjunk legalább az osztály szintjén, vagy ha szélesebb körű felmérésről van szó, akkor iskolai, megyei vagy országos szinten.

Két tesztet akkor tekintünk ekvivalensnek, ha a mérést a tesztekkel elvégezve a két teszt eredményei minden tanuló esetében megegyeznek. Ezt az ideális határesetet a gyakorlatban természetesen nem lehet elérni, esetleg csak többé-kevésbé megközelíteni. Az ekvivalencia biztosításához bonyolultabb matematikai-tesztelméleti megfontolásokat is felhasználhatunk, itt azonban csak az ekvivalencia egyszerűbb kérdéseivel foglalkozunk. A gyakorlatban általában két tesztet már akkor is ekvivalensnek tekinthetünk, ha a két teszttel egy populációt azonos módon lehet felmérni. Ahhoz, hogy a különböző tesztváltozatok ekvivalensek legyenek, többek között az kell, hogy mindegyik teszt megközelítően ugyanazt a tágabb tudás-

területet vizsgálja, a tesztekben hasonló arányban legyenek könnyű és nehéz feladatok.

Az ekvivalens tesztváltozatok szerkesztésekor a z e l k é s z í t e t t f e l a d a t o k a t a r á n y o - s a n o s z t j u k s z é t a k ü l ö n b ö z ő v á l - t o z a t o k b a . Az arányosság többféle szempont egyedejű figyelembevételét igényli. Szükség van arra, hogy a t a n - a n y a g e l e j é r ő l , k ö z e p é r ő l é s v é - g é r ő l e g y a r á n t k e r ü l j e n e k f e l a - d a t o k m i n d e g y i k t e s z t v á l t o z a t b a . Emellett a feladatoknak a b e n n ü k m e g - t e s t e s í t e t t t u d á s f o n t o s s á g á t é s a n e h é z s é g e t t e k i n t v e i s a r á n y o - s a n k e l l m e g j e l e n n i ü k a z e g y e s v á l t o z a t o k b a n .

Az ekvivalens tesztváltozatok készítése során másként járunk el a tudás ismeret-jellegű és képesség-jellegű elemeit mérő feladatokkal. Az ismeret-jellegű elemek mindig a maguk konkrétságában jelennek meg, egy adott tény vagy definíció mindig ugyanaz a tény vagy definíció lesz. Például a mohácsi csata évére többféleképpen rákérdezhetünk, többféle tesztfeladatot készíthetünk, a helyes megoldás mindig az 1526 lesz. Ezzel szemben a tudás képesség-jellegű összetevői sok, többé-kevésbé különböző konkrét tevékenységben megnyilvánulhatnak. Például a szöveges feladatok megoldásának tudását végtelenül sok pontosan azonos szerkezetű, de különböző szövegű és különböző számértékeket tartalmazó feladattal lehet megvizsgálni. Hasonlóképpen az írás- vagy olvasáskészség felmérésére is sokféle szöveget használhatunk, és joggal feltételezhetjük, hogy mindegyik nagy pontossággal ugyanazt méri. Feltételezhetjük, hogy ha egy tanuló egy fizika feladatot meg tud oldani, akkor meg tud oldani egy ettől csak számértékekben különböző másik feladatot is. Nem ennyire szükségszerű azonban az, hogy ha valaki ismer egy adott történelmi tényt, akkor egy másikat is ismerni fog. E különbségeket figyelembe véve

a képesség-jellegű tudásnál mód van arra, hogy pontosan (vagy legalábbis nagyon jó közelítéssel) azonosan mérő feladatokat készítsünk, az ismeret-jellegű tudáselemekkel viszont csak annyit tehetünk, hogy azokat arányosan szétosztjuk a különböző tesztváltozatok között.

Az ekvivalens tesztváltozatok készítése a felsorolt alapelvek figyelembe vételével történhet, végső soron azonban csak többszöri kipróbálással, a feladatok cserélésével, javításával lehet jó közelítéssel azonosan mérő tesztváltozatokat készíteni, és ezek ekvivalenciáját csak tapasztalati úton lehet bizonyítani.

5.3. A tesztek kipróbálása és standardizálása

A tesztek elkészítése nem ér véget azzal, hogy a feladatokat megfogalmazzuk, belőlük változatokat állítunk össze, ezeket oldalakra tördeljük, sokszorosításra alkalmas formában legépeljük. Szükség van arra is, hogy a gyakorlatban kipróbáljuk azokat, és a tapasztalatok alapján korrigáljuk az esetleges hibákat.

A kipróbálás és a korrekció során elvárható alaposság attól függ, hogy ki, milyen célokkal készítette a tesztek. A pedagógusok saját használatra szánt tesztjeiket ritkán próbálják ki, esetleg az ismételten alkalmazott tesztek az előző tapasztalatok alapján átdolgozzák. Ma már szinte minden középiskola, de sok általános iskola is rendelkezik akkora (vagy nagyobb) számítástechnikai kapacitással, mint amilyen-
nel tíz évvel ezelőtt csak a kiemelt kutatóintézetek rendel-

keztek. Nincs tehát akadálya annak, hogy a tanárok tesztjeiket kipróbálják és igényes elemzéseknek vessék alá.

Szélesebb körű felhasználásra szánt, szakértők által készített teszteknek mindenképpen át kell esniük a kipróbáláson, és az eredmények statisztikai elemzésén alapuló korrekción. Az igazán jó tesztek csak hosszabb fejlesztő folyamat eredményeként készülhetnek el. Ennek részleteivel a következő fejezetben foglalkozunk.

A tesztek kipróbálásán túl szükség van a végleges, tartós használatra szánt tesztek "bemérésére", különböző paramétereik megállapítására. Ennek egy sajátos formája a tesztek standardizálása, melynek során azt vizsgáljuk meg (felmérjük, vagy statisztikai eszközökkel becsüljük), hogy a tesztelni szándékozott személyek teljes köre (a teljes populáció) milyen teljesítményeket ért el a teszten.

A tesztek standardizálása a norma-orientált teszteknel hozzátartozik a pontozási rendszer kialakításához, hiszen ebben az esetben az egyes személyek teljesítményeit a populáció normáihoz viszonyítjuk. Például az intelligencia-teszteknel a populáció átlaga lesz a 100 IQ pont, a szórásnak pedig (általában) 15 pontot feleltetnek meg.

A tudásszintmérő teszteknel más a standardizálás funkciója. Ezeknel a teszteknel ki lehet alakítani a csoport normáitól független értékelő rendszert is. Ilyen megoldás lehet például az, hogy az eredményeket a maximális pontszám százalékában adjuk meg, tehát az egyes tanulók eredményeit a csoport eredményeitől függetlenül is értelmezhetjük. A standardizálásnak ebben az esetben az a szerepe, hogy a pedagógusok illetve a tanulók munkájuk értékeléséhez külső szempontokat kapjanak, saját eredményeiket összehasonlíthassák az országos eredményekkel.

A standardizáláshoz országos reprezentatív mintát kell választani. Figyelembe véve a szükséges pontosságot, erre a célra tesztváltozatónként elegendő 300 körüli tanuló. A megfelelő módon kiválasztott tanulók megoldják a teszteket, majd

eredményeik alapján kiszámítjuk a tesztek összpontszámának országos átlagát. Hasonlóképpen minden egyes itemre kiszámítjuk, hogy a tanulók hány százaléka oldotta meg helyesen. Ezeket az adatokat célszerű a tesztekkel együtt eljuttatni a pedagógusokhoz, így azok saját osztályuk, iskolájuk eredményeit az országos színvonalhoz viszonyíthatják.

A standardizálás lehetőséget adhat arra is, hogy a tanulók osztályozását országosan egységesítsük. Mivel a tudás-szintmérő tesztek elkészítése a kritérium-orientált szemléletmódon alapszik, vagyis a tesztek a tananyagot képezik le, lehetne az osztályozást úgy is egységesíteni, hogy például minden teszten a 85% feletti teljesítményre adunk ötöst, 70% felett négyest stb. Ilyen értékelő rendszert alkalmaznak a megtanító stratégiák, ahol csak azok a tanulók haladhatnak tovább, akik a tananyagot egy előre meghatározott szinten elsajátították. (Többnyire csak kétféle minősítés van: elsajátította, nem sajátította el.) A mai iskolákban azonban a tanulók többsége nem éri el a teljes elsajátítás szintjét, hanem egy részük nagyon is hiányos tudással halad tovább. Az elsajátítás különbségei az osztályzatok különbségeiben jelennek meg. Az egyes témakörök elsajátításában is nagy különbségek lehetnek: egyik témakört a tanulók többsége jól elsajátítja, a másikat kevésbé. Az értékelés gyakorlatában azonban nagy hagyománya van annak a szemléletmódnak, amelyik a tanulókat egymáshoz viszonyítva osztályozza, a jelesek, jók, közepesek stb. egymáshoz viszonyított aránya nagyjából mindig ugyanolyan, függetlenül attól, hogy könnyű vagy nehéz tananyagról van szó, hogy az elsajátítás szintje átlagosan jó vagy gyenge.

Az iskola létező mechanizmusaira, elvárására reagál a standard osztályzat. Ez a százalékpontoknak egy olyan osztályzattá alakítási rendszere, amely az országos eredményeket figyelembe véve úgy jelöli ki az egyes érdemjegyek pontszám-határait, hogy országos szinten mindig a tanulók egy meghatározott hányada kapjon ötöst, négyest stb.

A nehezebb teszteknel tehát a határok alacsonyabbra, a könnyebbeknél magasabbra kerülnek.

6. TESZTELEMZÉS, TESZTFEJLESZTÉS

Jó tesztek csak hosszabb fejlesztés eredményeként lehet készíteni. Egy széleskörű használatra szánt tesztet elterjesztése előtt legalább kétszer ki kell próbálni ahhoz, hogy megfelelő színvonalú mérőeszközzé váljon. Az oktatásban huzamosan használt tesztek pedig érdemes időről időre felülvizsgálni, átdolgozni, továbbfejleszteni. Sajnos nálunk még nem intézményesültek ennek az eljárásai, a létező példák is inkább kivételek. A tesztelemezés és tesztfejlesztés ma már sokféle matematikai, számítástechnikai eszközt használhat. A következőkben ezekből csak a legismertebb, személyi számítógéppel is megvalósítható eljárásokat ismertetjük.

6.1. A feladatok, itemek jóságának vizsgálata

A feladatok, itemek jóságának elemzésére, a hibás itemek kiszűrésére csak két módszert mutatunk be, egyet a klasszikus tesztelméletre és egyet a kritérium-orientált tesztelméletre alapozott eljárásokból.

Ha feltételezzük, hogy a tesztünk homogén, vagyis minden feladat hozzávetőlegesen ugyanazt a tudást méri, akkor elvárhatjuk, hogy a feladatok eredményei szorosan korreláljanak egymással. Ki kell tehát számítanunk a korrelációs mátrixot,

azaz minden feladatnak minden más feladattal vett korrelációját. Ha olyan feladatot találunk, amelynek a korrelációi többnyire alacsonyak, akkor arra gyanakodhatunk, hogy az adott feladat nem azt a tudást méri, amit a többi feladat. Például a megoldásában nem azok az ismeretek, készségek dominálnak, amit az adott témakörben a tanulók megszereztek, hanem a másutt, korábban megszerzett tudás.

Szélsőséges esetekben előfordulhatnak a korrelációs mátrixban negatív értékek is. Ha sok a negatív korreláció, akkor szinte biztosra vehetjük, hogy az adott feladattal valami baj van. Ez ugyanis azt jelenti, hogy a jó tanulók (akik a feladatok többségét jól oldották meg), ezt a feladatot elrontották. Ilyen esetekben rendszerint rövid elemzéssel ki lehet deríteni, hogy a feladat rosszul van megfogalmazva, vagy a javítókulccsal van valami baj.

Ugyanezt az információt tömörebben tartalmazza a feladat-tesztösszpontszám korreláció. Ha egy feladat a tesztösszpontszámával alacsonyan korrelál, az jelzi, hogy a megoldásához szükséges tudáselem idegen a témakör egészétől, a negatív korreláció pedig a feladat hibájára utal. Hasonlóan elemezhetjük a feladatok részeit, az itemeket is.

A feladatok korrelációs mátrixa és a feladat-összpontszám korrelációkban rejlő információk között vannak finom különbségek, amelyeket ugyancsak felhasználhatunk a hibásnak tűnő feladatok vizsgálatánál. A korrelációs mátrix egésze, sok korrelációs együttható együttesen ugyanis kialakít egy képet arról, hogy milyen magas a korrelációk többsége. Ha a korrelációs együtthatók többnyire magasak, az azt jelzi, hogy a teszt egésze homogén tudást vizsgál. Ebben az esetben egy feladat alacsonyabb együtthatói már jelezhetnek valamit. Ha azonban a korrelációk többsége alacsony, akkor a teszt inhomogén tudást vizsgál, és csak a többségnél sokkal kisebb, vagy kifejezetten negatív korrelációk utalnak hibás feladatra.

A tudásszintmérő teszteként soha nem várhatjuk el, hogy

teljesen homogének legyenek, vagyis minden itemjük ugyanazt a tudást mérje, ezért a klasszikus tesztelméletre alapozott elemzések csak korlátozott érvényűek lehetnek. Nem lenne helyes a feladatokat addig változtatgatni, amíg végül csupa magas korrelációt kapunk, ezzel ugyanis esetleg megfosztjuk a tesztet minden, az adott tananyagrészre jellemző specifikumától, és csak olyan feladatok maradnak benne, amelyek megoldásához már csak valamelyik általános képességre van szükség.

Más jellegű technikákat javasol az itemek elemzésére a kritérium-orientált tesztelmélet. Az itemek jóságának megítélésénél a hangsúly inkább azon van, hogy mire akarjuk a tesztet használni, és arra a célra megfelel-e az adott item.

Mivel a tudásszintmérő tesztek tipikus alkalmazása az oktatási folyamat hatékonyabb irányítása, a kritérium-orientált tesztelmélet az elemzések többségét erre a célra dolgozta ki. Például a tesztelés célja az lehet, hogy megállapítsuk, a tanulók mennyit tanultak egy adott időszakban. Nem az tehát a célunk, hogy felmérjük, általában mit tudnak a szóbanforgó kérdésekről, hanem arra vagyunk kíváncsiak, hogy az adott tanítási periódus mennyire volt számukra hatékony. Ilyenkor azokat az itemeket akarjuk a tesztből kiszűrni, amelyek nem ilyen tudást mérnek. Erre a célra felhasználhatjuk az oktatással szembeni érzékenység mutatóját.

Az itemek oktatással szembeni érzékenysége azt jelenti, hogy az adott item megoldásának esélyét hogyan befolyasolja az adott tanítási-tanulási periódus. Nyilvánvalóan azok az itemek, amelyeket a tanulók már az oktatás előtt is jól meg tudnak oldani, nem az adott időszakban megszerzett tudást mérik. Hasonlóképpen azok az itemek sem, amelyeket a tanulók az adott oktatás hatására sem tudnak megoldani.

E gondolatmenet felhasználásával megkonstruálhatjuk az oktatással szembeni érzékenység mutatóját úgy, hogy képezzük az item megoldásának oktatás előtti és oktatás utáni átlagát. E mutató kiszámításához a tesztet kétszer kell a tanulóknak megírniuk: az adott témakör tanítása előtt és után. Ekkor

1-gyel értékelve minden jó és 0-val minden rossz megoldást, az itemek átlageredményei 0 és 1 közé fognak esni. Minden egyes itemre képezzük az oktatás előtti és az oktatás utáni eredmények különbségét. Az így kapott mutatók -1 és +1 között változhatnak. Minél nagyobb ez az érték, annál inkább az adott oktatás hatására elsajátított tudást méri az item. Legmagasabb értéke 1 lehet, ebben az esetben az oktatás megkezdése előtt senki sem tudta a feladatot megoldani, utána viszont mindenki. Akkor adódik 0, ha az oktatásnak nincs hatása az eredményekre, tehát előtte és utána pontosan ugyanannyian oldják meg jól az itemet. A negatív érték azt jelzi, hogy az oktatás után kevesebben oldották meg az itemet, mint előtte. Ezt (hacsak nem tételezzük fel az oktatás butító hatását) már mindig valamilyen hiba jelzésének kell tekintenünk.

Ha el akarjuk kerülni azt, hogy ugyanazzal a tanulócsoporthal kétszer megírassuk a tesztet, hasonló megfontolások alapján végezhetünk elemzéseket egy-egy, a szóbanforgó oktatásban részt vett és részt nem vett csoporttal. Bár gyakorlatilag nehéz lehet két egyenértékű, egy oktatott és egy nem oktatott csoportot találni, az oktatott és a nem oktatott csoport eredményeinek különbségei is jól jellemzik azt, hogy mely itemek mérik az oktatás hatására kialakult tudást.

Ez utóbbi elemzésekre a klasszikus megoldásokkal teljesen ellentétes hatású tesztfejlesztést alapozhatunk. Ha azokat az itemeket hagyjuk ki a tesztből, illetve javítjuk ki, amelyeknek alacsony az oktatással szembeni érzékenységet jellemző indexük, akkor a tesztünk egyre inkább azt a speciális tudást méri, amit az adott oktatási szakasz eredményeként szereztek meg a tanulók.

6.2. A reliabilitásmutatók kiszámítása

A tesztfejlesztés egyik célja a tesztek megbízhatóságának, reliabilitásának javítása. Ezért a tesztfejlesztés során annak minden fázisában ki kell számítani a reliabilitásmutatókat, hogy képet kapjunk a tesztfeladatokon végzett változtatások hatásáról, hogy lássuk, mennyit kell még a feladatokon csiszolni, hogy megbízható tesztet kapjunk. Ezen kívül a tesztek felhasználóját is tájékoztatnunk kell arról, hogy a teszt milyen jól mér, mennyire bízhatunk meg az általa szolgáltatott eredményekben.

A reliabilitásmutatók kiszámítására sokféle formula áll rendelkezésünkre. Mint korábban már kifejtettük, ezek mindegyike alsó becslést ad a teszt reliabilitásának valódi értékére. A különböző formulák különböző értékeket adnak a reliabilitásra, de bármelyik formulával is számolunk, a reliabilitás valódi értéke mindegyiknél csak nagyobb vagy vele megegyező lehet.

A formulák között azt mérlegelve válogathatunk, hogy milyen adatok állnak rendelkezésünkre, és milyen eszközzel végezzük a számításokat. Kézi vagy zsebszámológéppel dolgozva választhatunk kevesebb munkát igénylő, de gyengébb becslést adó formulát. Számítógépes adatfeldolgozás esetén érdemes számításigényesebb, de jobb becslést adó formulát használni.

Ha a tesztfelezéses módszerrel akarjuk a teszt reliabilitását meghatározni, külön számítjuk ki a teszt páros és páratlan sorszámú itemjeinek az összpontszámát. Majd kiszámítjuk a két érték közötti korrelációs együtthatót. Ezt jelöljük r -rel. Az r felhasználásával a Sperman-Brown formula segítségével számíthatjuk ki a reliabilitásmutatót (ezt r_{tt} -vel jelöljük).

$$r_{tt} = \frac{2r}{1+r}$$

A tesztek belső konzisztenciájának jellemzésére szolgáló legszemléletesebb mutató a Cronbach-féle alfa koefficiens. Ez az itemek és a teszt varianciáiból indul ki, ha a szórások rendelkezésünkre állnak, kiszámítása viszonylag egyszerű.

$$\alpha = \left(\frac{n}{n-1} \right) \cdot \left(1 - \frac{\sum s_i^2}{s_t^2} \right)$$

ahol n az itemek számát, s_i az itemek szórását, s_t a tesztösszpontszám szórását jelöli.

Ha rendelkezésünkre állnak az itemek átlagai és az item-tesztösszpontszám korrelációk, használhatjuk a Gulliksen formulát:

$$r_{tt} = \left(\frac{n}{n-1} \right) \cdot \left(1 - \frac{\sum pq}{\left(\sum r_{it} \sqrt{pq} \right)^2} \right)$$

Itt p az itemek megoldásának arányát jelöli, 1-gyel a jó, 0-val a rossz megoldásokat értékelve ez megegyezik az átlaggal. A $q=1-p$, az r_{it} pedig az item-tesztösszpontszám korreláció.

A Kuder-Richardson 20 formula használatához elegendő az itemek (súlyozás nélküli) átlagát és a tesztösszpontszám szórását ismerni. Az előző jelöléseket felhasználva:

$$r_{tt} = \left(\frac{n}{n-1} \right) \cdot \left(1 - \frac{\sum pq}{s_t^2} \right)$$

A tesztek reliabilitását a folyamatos fejlesztő munkával igen nagy mértékben javítani lehet. Már néhány hibás item kijavításával, kicserélésével is jelentős javulást lehet elérni.

A gyakorlatban használt egyszerűbb tesztek¹ől megkövetelhetjük, hogy reliabilitásmutatójuk legalább a 0,8 legyen. A hosszas fejlesztő munka eredményeként elkészült tesztek reliabilitása pedig elérheti 0,95 - 0,98 értékeket is.

I R O D A L O M

Az alábbi jegyzék az egyes témakörök további tanulmányozásához felhasználható fontosabb magyar, angol és német nyelvű munkákat tartalmazza.

A pedagógiai értékeléshez:

Báthory Zoltán: Tanítás és tanulás

Tankönyvkiadó, Budapest, 1985

Nagy József: Köznevelés és rendszerszemlélet

Országos Oktatástechnikai Központ, Veszprém, 1979

A klasszikus tesztelmélethez:

Guilford, J.P.: Fundamental Statistics in Psychology and Education

McGraw Hill, New York, 1965

Lienert, G. A.: Testaufbau und Testanalyse

Julius Beltz, Weinheim und Berlin, 1967

Lord, F.M. and Novick, M.R.: Statistical Theories of Mental Test Scores

Addison-Wesley, Reading, Mass., 1968

A kritériumorientált teszteléshez:

Berk, R.A. (ed.): Criterion-referenced measurement: The state of the art

The Johns Hopkins Press Ltd., London, 1980

Csapó Benő: A kritériumorientált értékelés

Magyar Pedagógia, 1987, Megjelenés alatt

Fricke, R.: Kriteriumsorientierte Leistungsmessung

- Verlag W. Kohlhammer, Stuttgart, 1974
- Popham, W.J.: Criterion-referenced measurement
Prentice Hall, Englewood Cliffs, N. J., 1978
- Srittmatter, P. (Hrsg.): Lernzielorientierte Leistungsmessung
Beltz, Weinheim, 1973

Az újabb tesztelméletekről:

- Horváth György: Tesztelmélet: problémák és perspektívák
Pszichológia, 1985. 1. sz. 53-78. o.

A tananyag elemzéséhez:

- Bloom, B.S.: Taxonomy of Educational Objectives: Cognitive
Domain
McKay, New York, 1956
- Kádárné Fülöp Judit - Joó András: Beszámoló a strukturális
elemzés pedagógiai alkalmazásának néhány módszeréről.
OPI dokumentumok 8. OPI, Budapest, 1977.
- Nagy József: A tudástechnológia elméleti alapjai
Országos Oktatástechnikai Központ, Veszprém, 1985
- Orosz Sándor: A tananyag elemzése
Országos Oktatástechnikai Központ, Budapest, 1977

A teszt szerkesztéshez:

- Ágoston György - Nagy József - Orosz Sándor: Mérések módsze-
rek a pedagógiában
Tankönyvkiadó, Budapest, 1979.
- Bloom, B.S., Hastings, J.I., Madaus, G.F.: Handbook of
Formative and Summative Evaluation of Student Learning
McGraw Hill Co. 1971
- Joó András: A feladat készítés kérdései
Országos Oktatástechnikai Központ, Veszprém, 1980
- Nagy József: A témazáró tudásszintmérés gyakorlati

kérdései

Tankönyvkiadó, Budapest 1972

Thorndike, R.L. (ed.): Educational Measurement

American Council on Education, Washington, 1971

Példák a pedagógiai tesztekre:

Báthory Zoltán: 7 standardizált tantárgyteszt

Országos Pedagógiai Intézet, Budapest, 1973

Csapó Benő - Varsányi Zoltán: A rajzkészség fejlettségének vizsgálata középiskolai tanulóknál

Acta Paedagogica, Ser. Spec., Szeged, 1985.

Csáki Imre - Nagy József: Alsó tagozatos szöveges feladatbank

Acta Paedagogica, Ser.Spec., Szeged, 1976

Kozéki Béla: Személyiségfejlesztés az iskolában

Békés Megyei Pedagógiai Intézet, Békéscsaba, 1984

Nagy József: Az elemi számolási készségek

Tankönyvkiadó, Budapest, 1971

Nagy József: Alapműveleti számolási készségek

Acta Paedagogica, Ser. Spec., Szeged, 1973

Orosz Sándor: A helyesírás fejlődése

Tankönyvkiadó, Budapest, 1974

Orosz Sándor: A fogalmazástechnika mérésmethodikai problémái

Tankönyvkiadó, Budapest, 1974

Standardizált Témazáró Tesztek

17 kötetnyi teszt az általános iskola felsőtagozatának tananyagához

Acta Univ. Szeg. de A.J. nom. Sectio Paed. Ser.Spec.,
Szeged, 1972-1976